NRL Report 9051
(Revised)

AD-A207 851

Convergence Rate of a Gram-Schmidt Canceller

KARL GERLACH AND FRANK F. KRETSCHMER, JR.

Target Characteristics Branch
Radar Division

April 20, 1989

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No 0704 0188

1a REPORT SECURITY CLASSIFICATION	UNCLASSIFIED
1b RESTRICTIVE MARKINGS	
2a SECURITY CLASSIFICATION AUTHORITY	
2b DECLASSIFICATION / DOWNGRADING SCHEDULE	
3 DISTRIBUTION / AVAILABILITY OF REPORT	Approved for public release; distribution unlimited.
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	NRL Report 9051 (Revised)
5 MONITORING ORGANIZATION REPORT NUMBER(S)	
6a NAME OF PERFORMING ORGANIZATION	Naval Research Laboratory
6b OFFICE SYMBOL (If applicable)	Code 5340
7a NAME OF MONITORING ORGANIZATION	Office of the Chief of Naval Research
6c ADDRESS (City, State, and ZIP Code)	Washington, DC 20375-5000
7b ADDRESS (City, State, and ZIP Code)	Arlington, VA 22217-5000
8a NAME OF FUNDING / SPONSORING ORGANIZATION	Space and Naval Warfare Systems Command
8b OFFICE SYMBOL (If applicable)	
9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8c ADDRESS (City, State, and ZIP Code)	Washington, DC 20363-5100

10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
(see Page ii)	(see Page ii)	(see Page ii)	(see Page ii)

11 TITLE (Include Security Classification)

Convergence Rate of a Gram-Schmidt Canceller

12 PERSONAL AUTHOR(S)

Gerlach, Karl and Kretschmer, F.F., Jr.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Interim	FROM ______ TO ______	1989 April 20	42

16 SUPPLEMENTARY NOTATION

17 COSATI CODES		
FIELD	GROUP	SUB-GROUP

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

Adaptive filter. ECCM
Radar Adaptive cancellation.
Gram-Schmidt

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The open-loop Gram-Schmidt (GS) canceller is shown to be numerically identical with the Sampled Matrix Inversion (SMI) algorithm in the transient state if infinite numerical accuracy is assumed. Three forms of the GS canceller are discussed and analyzed—concurrent, nonconcurrent, and sliding window processing. Convergence results for concurrent and nonconcurrent SMI cancellers have been obtained in the past by Reed, Mallet, and Brennan under the assumption that the inputs are Gaussian. In this report many of those results are reproduced by using the GS structures as an analysis tool. In addition, new results are obtained for when the input noises are not Gaussian. Furthermore, it is shown that the sliding window GS canceller has the same convergence properties as the concurrent GS canceller. The deleterious effect of "over matching the degrees of freedom" is discussed.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT	☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS
21 ABSTRACT SECURITY CLASSIFICATION	UNCLASSIFIED
22a NAME OF RESPONSIBLE INDIVIDUAL	Karl Gerlach
22b TELEPHONE (Include Area Code)	(202) 767-3599
22c OFFICE SYMBOL	Code 5340

DD Form 1473, JUN 86 Previous editions are obsolete SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
62712N	RS12-131-001	XF12-141-100	DN480-549 DN380-101

CONTENTS

I. INTRODUCTION 1

II. THE GS CANCELLER 2

III. OUTPUT MEASURES 3

IV. SMI AND GS CANCELLER EQUIVALENCE 8

V. INVARIANT TRANSFORMS AND GS THEOREMS 12

VI. THE TWO-INPUT GS CANCELLER 17

VII. GENERAL MOMENT THEOREM FOR GS CANCELLERS 21

VIII. CONVERGENCE RESULTS 24

IX. THE CONCURRENT GS CANCELLER 25

X. LOWER BOUND 30

XI. SLIDING WINDOW GS CANCELLER CONVERGENCE 32

XII. OVERMATCHING DEGREES OF FREEDOM 33

XIII. SUMMARY 33

XIV. REFERENCES 34

APPENDIX A — Probability Density Function Associated with a Two-Input GS Canceller 35

APPENDIX B — First Moment Bound 37

CONVERGENCE RATE OF A GRAM-SCHMIDT CANCELLER

I. INTRODUCTION

The open-loop Gram-Schmidt (GS) technique for adaptive cancellation [1-6] has been shown to yield superior performance simultaneously in arithmetic efficiency, stability, and convergence times over other adaptive algorithms. Arithmetic efficiency results from using systolic processing architectures that take advantage of the GS structure. In addition, the stability of the GS algorithm is enhanced because it does not require the direct calculation of an inverse covariance matrix as does the Sampled Matrix Inversion (SMI) algorithm [7]. Also, the GS canceller algorithm is very suitable for a nonstationary noise environment because the adaptive weights can be updated in a numerically efficient manner, by using "sliding window" techniques on the input data instead of "batch" or "block" processing. Two types of batch processing techniques are concurrent processing and nonconcurrent processing. For concurrent processing, the adaptive weights are calculated from an input data set and reapplied to the same input data set. For nonconcurrent processing, the weights are applied to a different data set.

The optimal weights associated with an adaptive canceller are generally not known a priori and thus must be estimated by using finite averaging. Because of the use of estimated weights, suboptimal canceller performance results. Reed, Mallet, and Brennan [7,8] quantified this performance for the SMI algorithm in the transient state under certain input conditions, one of these being that the input noise must be Gaussian. They mathematically demonstrated that the SMI canceller has relatively fast convergence characteristics and also that the convergence is independent of the input covariance matrix.

In this report, we show that the GS canceller algorithm is numerically identical with the SMI algorithm in the transient state if infinite numerical accuracy is assumed (Section IV). By transient state, we mean that a finite number of time-coincident samples per channel are used to obtain an estimate of the optimal weights. Thus the convergence rate or any other measures of effectiveness of the two algorithms in the transient state are identical. In addition, we reproduce many of the results of Refs. 7 and 8 by using the GS canceller structures as an analysis tool (Sections V to VIII). Also, new results are generated for the case when the input noises are not necessarily Gaussian (Sections IX and X). Furthermore, various results are presented for the three forms of the GS canceller (concurrent processor, nonconcurrent processor, and sliding window processor). In particular, it is shown that the sliding window GS canceller is convergent, equivalent in the transient state with the concurrent GS canceller (Section XI). Section XII discusses the deleterious effect of "overmatching the degrees of freedom."

Note that the analysis presented in this report pertains to the adaptive processor in the "sidelobe canceller" (SLC) configuration, where the desired signal is assumed to be present only in the main channel and auxiliary channels are used to cancel correlated noises in the main channel. However, as is shown in Ref. 7, any nonconstrained, linear adaptive array processor can be transformed into an SLC configuration without changing the convergence properties. Hence, the results of this report apply to any nonconstrained linear adaptive array processor.

Manuscript approved March 9, 1989.

II. THE GS CANCELLER

Consider the general N-input GS canceller structure as seen in Fig. 1(a). Let $x_M(t)$, $x_1(t)$, . . . , $x_{N-1}(t)$ represent the complex data in the 0th, 1st, . . . , $N-1$th channels, respectively. We call the leftmost input $x_M(t)$ the main channel, and we call the remaining $N-1$ inputs the auxiliary channels. The main channel's signal consists of a desired signal plus additive noise. The noise consists of internal noise plus external noise. Cancellation of the signals from external interfering sources relies on the correlation of simultaneously received signals in the main and auxiliary channels. The internal noises on each channel are assumed uncorrelated between channels. The canceller operates so as to decorrelate the auxiliary inputs one at a time from the other inputs by use of the basic two-input GS processor as is shown in Fig. 1(b). For example, Fig. 1(a) shows that $x_{N-1}(t)$ is decorrelated with $x_M(t)$, $x_1(t)$, . . . , $x_{N-2}(t)$ in the first level of decomposition. Next, the output channel that results from decorrelating $x_{N-1}(t)$ with $x_{N-2}(t)$ is decorrelated with the other outputs of the first-level GSs. The decomposition proceeds until a final output channel is generated. If the decorrelation weights in each of the two-input GSs are computed from an infinite number of input samples, this output channel is totally decorrelated with the input: $x_1(t)$, $x_2(t)$, . . . , $x_{N-1}(t)$.

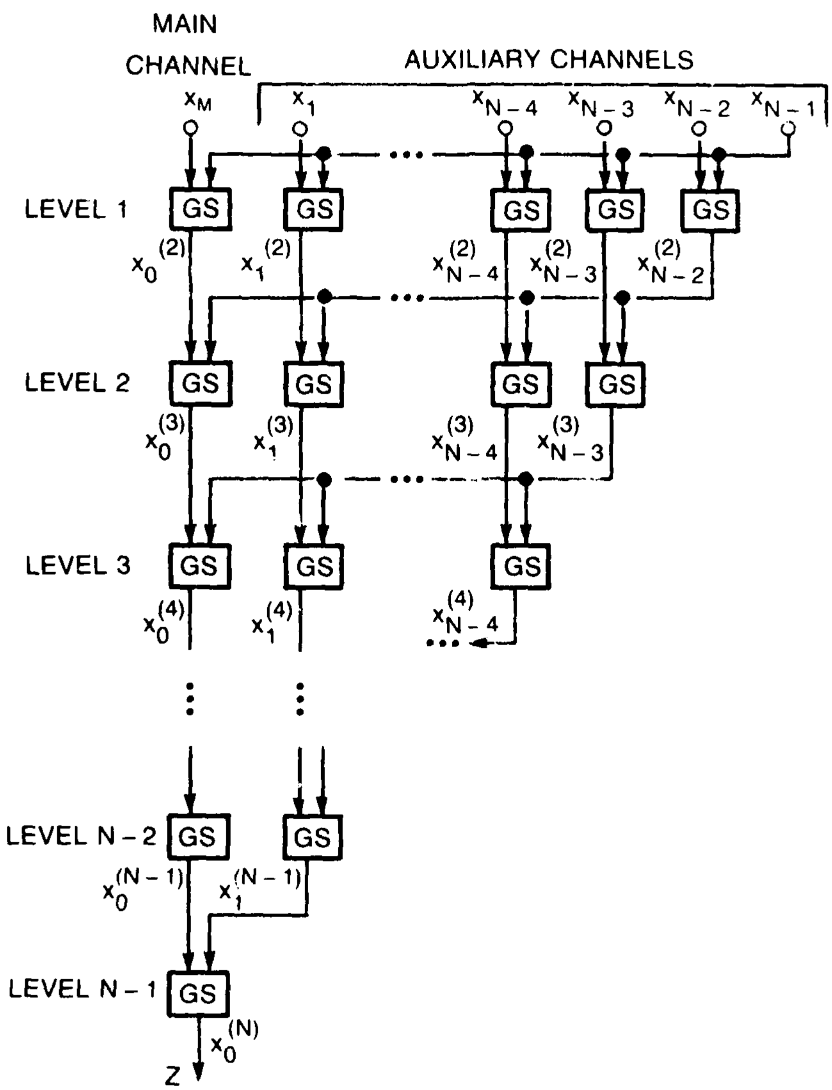

Fig. 1(a) — GS structure

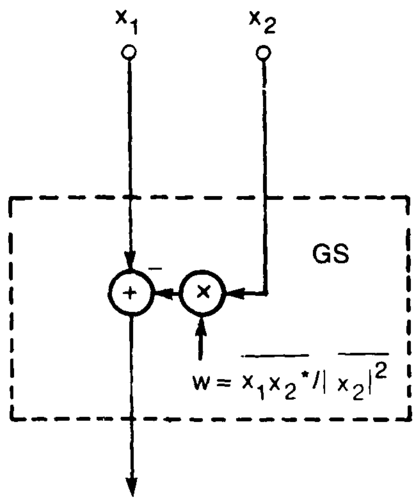

Fig. 1(b) — Basic two-input GS canceller

If there is not an infinite number of input samples, the decorrelation weights associated with each two-input GS canceller are estimated by using finite averaging. In this section we discuss two methods of processing data through the GS canceller. The first is called concurrent processing whereby the weights are estimated from a block of input data and applied back onto the same input data set. The second method is called nonconcurrent processing whereby the weights are estimated from a block of input data and applied to subsequent or previous input data. Inherent in both techniques is the "block processing" of data. (In Section XI we consider another form of the GS canceller; it uses sliding window or systolic techniques.) Reference 8 shows that the average output noise power residue can vary quite differently, depending on whether concurrent or nonconcurrent processing is used.

We now briefly describe the concurrent and nonconcurrent GS canceller. For the concurrent canceller, let $x_n^{(m)}$ represent the time-coincident outputs of the two-input GSs on the $(m - 1)$th level. Then outputs of the two-input GSs at the mth level are given by

$$x_n^{(m+1)} = x_n^{(m)} - w_n^{(m)} x_{N-m}^{(m)}, \quad \begin{array}{l} n = 0, 1, \ldots, N - m - 1, \\ m = 1, 2, \ldots, N - 1. \end{array} \tag{2.1}$$

Note that $x_0^{(1)} = x_M$ and $x_n^{(1)} = x_n$, $n = 1, 2, \ldots N - 1$. The weight $w_n^{(m)}$, seen in Eq. (2.1), is computed so as to decorrelate $x_n^{(m)}$ with $x_{N-m}^{(m)}$. For K input samples per channel, this weight is estimated as

$$w_n^{(m)} = \frac{\sum_{k=1}^{K} x_{N-m}^{(m)*}(k) x_n^{(m)}(k)}{\sum_{k=1}^{K} |x_{N-m}^{(m)}(k)|^2}, \tag{2.2}$$

where * denotes the complex conjugate and $|\cdot|$ is the magnitude. Here k indexes the time-coincident sampled data.

For the nonconcurrent canceller, let $X_n^{(m)}$ represent the outputs of the two-input GSs on the $(m - 1)$th level. Then the outputs of the two-input GSs at the mth level are given by

$$X_n^{(m+1)} = X_n^{(m)} - w_n^{(m)} X_{N-m}^{(m)}, \quad \begin{array}{l} n = 0, 1, \ldots, N - m - 1 \\ m = 1, 2, \ldots, N - 1, \end{array} \tag{2.3}$$

where $X_0^{(1)} = X_M$, $X_n^{(1)} = X_n$, $n = 1, 2, \ldots N - 1$, and $w_n^{(m)}$ is calculated by the use of Eq. (2.2); i.e., these weights are computed from a block of data that does not include X_n.

Let x_0 and X_0 represent the additive noises in the main channel for concurrent and nonconcurrent processing, respectively. For this development, unless otherwise noted, we make the following assumptions:

1. The $x_0, x_1, \ldots, x_{N-1}$ and $X_0, X_1, \ldots, X_{N-1}$ are identically distributed Gaussian complex random variables (r.v.).

2. These same r.v.'s are samples from stationary processes with zero mean and equal variance.

3. For $k_1 \neq k_2$, $x_n(k_1)$ is independent of $x_n(k_2)$ and $X_n(k_1)$ is independent of $X_n(k_2)$.

4. $x_{n_1}(k_1)$ is independent of $X_{n_2}(k_2)$ for all k_1, k_2, n_1, n_2.

5. The desired signals are not present during weight computation for nonconcurrent processing.

6. The desired signals are not present in the auxiliary channels.

Note that in Sections IX and X, we remove the assumption that the r.v.'s are Gaussian.

The following definition is used often in the upcoming development. A normalized L-length multivariate complex circular Gaussian vector has L elements, each of which has real and imaginary parts that are independent Gaussian r.v.'s with 0 mean and variance equal to 1/2 (note the magnitude variance is one). In addition, the L elements are independent of one another.

III. OUTPUT MEASURES

The N-input GS canceller structures for concurrent and nonconcurrent processing are simplified by the representations as seen in Fig. 2(a). There are $.5N\ (N - 1)$ weights computed in the GS structure. We call these weights the GS interior weights. The notation $GS_{K,N}$ indicates that an N-input GS structure uses K samples from each channel to compute the GS interior weights in the GS structure. Note that for the nonconcurrent structure the weights are computed from the $x_0, x_1, \ldots, x_{N-1}$ data block and applied to $X_0, X_1, \ldots, X_{N-1}$. The 0th channel (or the far left channel in Fig. 2(a)) is always designated as the main channel, and the others are called auxiliary channels (or AUXs). The output of the concurrent (weighting) processor is denoted by z_{cw}, and the output of the nonconcurrent processor is denoted by Z_{nw}. We also represent the GS structure as shown in Fig. 2(b), where the N orthogonal outputs are displayed.

As previously mentioned, input signals in the main channel consist of a desired signal s plus noise x_0. For nonconcurrent processing, it is assumed that the desired signal passes from input to output unperturbed. However, for concurrent processing the presence of the signal in the main channel causes signal cancellation through the GS canceller as reported in Ref. 8. Because of linearity, the $GS_{K,N}$ canceller can be decomposed as shown in Fig. 3 (see Theorem 2, Section IV). The left-hand $GS_{K,N}$ canceller seen in this figure has only the desired signal in the main channel, and the

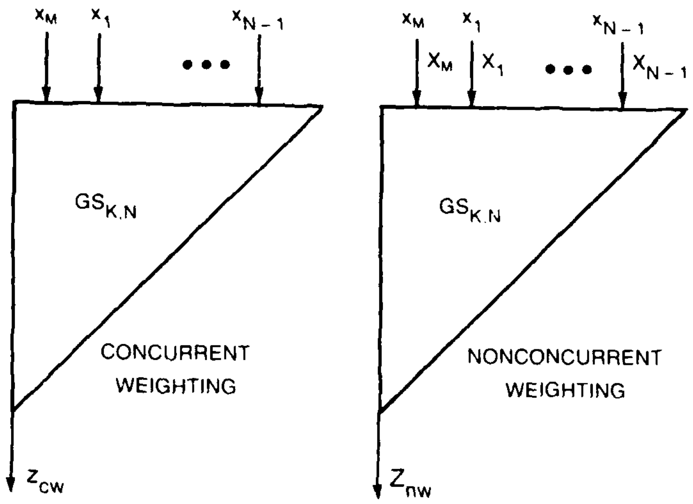

Fig. 2(a) — Representations of concurrent and nonconcurrent weighting of GS cancellers

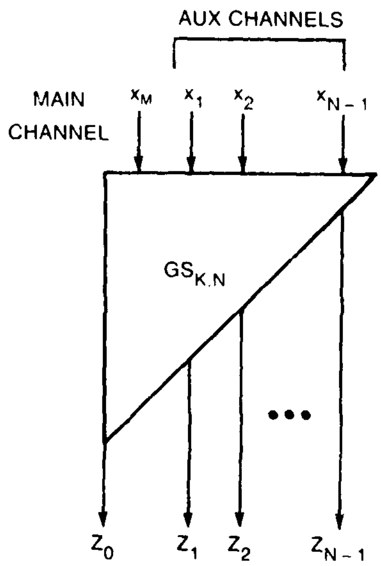

Fig. 2(b) — GS representation with N output channels

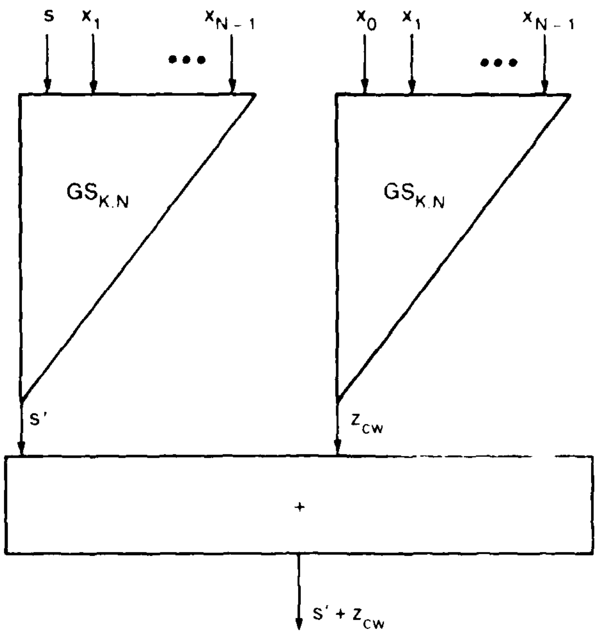

Fig. 3 — Decomposition of signal and noise for concurrent processing

right-hand $GS_{K,N}$ has only noise x_0 in the main channel. Note that the interior weights of each $GS_{K,N}$ are not identical because of the different main channel input in each (actually only the weights along the main channel path of each differ). As will be seen, we use this decomposition when defining the output noise-to-signal power ratio (NSR).

It can be shown that for any set of GS interior weights that are estimated and applied to the auxiliary input channels there is an equivalent linear weighting of the input auxiliary channels. We denote this equivalent linear weighting by the $(N-1)$-length vector $\hat{\mathbf{w}}_a$, where

$$\hat{\mathbf{w}}_a = (\hat{w}_1, \ldots, \hat{w}_{N-1})^T. \tag{3.1}$$

Thus, the outputs of the GS processed main channel are identical to the outputs of a main channel derived by subtracting the linear weighted auxiliary channels from the main channel input. With respect to the decomposition configuration seen in Fig. 3, we see that $\hat{\mathbf{w}}_a$ is actually the sum of two $(N-1)$-length weighting vectors $\hat{\mathbf{w}}_{a,s}$ and $\hat{\mathbf{w}}_{a,n}$, where $\hat{\mathbf{w}}_{a,s}$ is the auxiliary linear weighting vector associated with only the desired signal in the main channel and $\hat{\mathbf{w}}_{a,n}$ is the auxiliary linear weighting vector associated with only noise x_0 in the main channel. Note that as $K \to \infty$, $\hat{\mathbf{w}}_{a,s} \to \mathbf{0}$; also for nonconcurrent processing, $\hat{\mathbf{w}}_a = \hat{\mathbf{w}}_{a,n}$. For the GS canceller the weighting on the main channel is constrained to be one.

Let $\sigma^2_{\min}$ be the steady state $(K \to \infty)$ output noise power residue and let $\mathrm{SNR}_{\mathrm{opt}}$ be the steady state output signal-to-noise ratio. Note that the $\sigma^2_{\min}$ is identical for both concurrent and nonconcurrent processing as is $\mathrm{SNR}_{\mathrm{opt}}$. Here

R_a is the steady state $(N-1) \times (N-1)$ input noise convariance matrix of the auxiliary channels,

$\hat{R}_a$ is the estimated auxiliary input covariance matrix using $x_1, x_2, \ldots, x_{N-1}$ data (K samples per input channel),

$\hat{R}_X$ is the estimated auxiliary input noise covariance matrix using $X_1, X_2, \ldots, X_{N-1}$ (note, no desired signal assumed in this calculation),

$\hat{\sigma}^2_{nw}$ is the transient output noise power associated with nonconcurrent weighting normalized by dividing by $\sigma^2_{\min}$,

$\overline{\hat{\sigma}^2_{nw}}$ is the transient output noise power associated with nonconcurrent weighting normalized to $\sigma^2_{\min}$ and averaged over $X_0, X_1, \ldots, X_{N-1}$,

$\hat{\mathrm{SNR}}_{nw}$ is the transient output ratio SNR associated with nonconcurrent weighting normalized by dividing by $\mathrm{SNR}_{\mathrm{opt}}$ and averaged over $X_0, X_1, \ldots, X_{N-1}$,

$|s'|^2$ is the transient output signal power associated with concurrent weighting normalized by dividing by the input desired signal power $|s|^2$,

$\hat{\sigma}^2_{cw}$ is the transient output noise power associated with concurrent weighting normalized by dividing by $\sigma^2_{\min}$, and

$\hat{\mathrm{NSR}}_{cw}$ is the transient output noise-to-signal power ratio associated with concurrent weighting normalized by dividing by $\mathrm{SNR}_{\mathrm{opt}}$.

Note that the last eight quantities defined are r.v.'s. $\hat{\mathrm{NSR}}_{cw}$ is defined as a noise-to-signal ratio because it is easier to obtain an analytical result pertaining to this quantity as opposed to the SNR.

By using the above definitions, it can be shown that

$$\hat{\sigma}^2_{nw} = \frac{|Z_{nw}|^2}{\sigma^2_{\min}} = |x_0|^2 - \hat{\mathbf{w}}^t_a \hat{R}_X \hat{\mathbf{w}}_a \tag{3.2a}$$

and

$$\overline{\hat{\sigma}^2_{nw}} = E_X\{\hat{\sigma}^2_{nw}\} = \frac{E\{|x_0|^2\} - \hat{\mathbf{w}}^t_a R_a \hat{\mathbf{w}}_a}{\sigma^2_{\min}}, \tag{3.2b}$$

where $E\{\ \}$ denotes the expected value and $E_X\{\cdot\}$ denotes that the expectation is taken over the r.v.'s $X_0, X_1, \ldots X_{N-1}$. Furthermore,

$$\hat{\mathrm{SNR}}_{nw} = \frac{(1/\overline{\hat{\sigma}^2_{nw}})}{\mathrm{SNR}_{opt}}, \tag{3.3}$$

and

$$\hat{\sigma}^2_{cw} = \frac{|z_{cw}|^2}{\sigma^2_{\min}} = \frac{|x_0|^2 - \hat{\mathbf{w}}^t_{a,n} \hat{R}_a \hat{\mathbf{w}}_{a,n}}{\sigma^2_{\min}}, \tag{3.4}$$

$$\hat{\mathrm{NSR}}^2_{cw} = \frac{\hat{\sigma}^2_{cw}}{|s'|^2}. \tag{3.5}$$

We define

$$\sigma_{nw}^2(K,N) = E\{\hat{\sigma}_{nw}^2\} = E\{\overline{\hat{\sigma}_{nw}^2}\} \tag{3.6}$$

$$\mathrm{SNR}_{nw}(K,N) = E\{\hat{\mathrm{SNR}}_{nw}\}, \tag{3.7}$$

$$\sigma_{cw}^2\ (K,N) = E\{\hat{\sigma}_{cw}^2\}, \tag{3.8}$$

$$s_{cw}\ (K,N) = E\{|s'|^2\}, \tag{3.9}$$

and

$$\mathrm{NSR}_{cw}\ (K,N) = E\{\hat{\mathrm{NSR}}_{cw}\}. \tag{3.10}$$

Equations (3.6) to (3.10) are the first moments or average transient values of the previously defined output measures of the GS cancellers. These output measures are commonly used to grade the convergence performance of the SMI canceller.

In addition, the ith moment of $\hat{\mathrm{SNR}}_{nw}$ is defined as

$$\mathrm{SNR}_{nw}^{(i)}\ (K,N) = E\{(\hat{\mathrm{SNR}}_{nw})^i\}, \tag{3.11}$$

and the ith moment of $|s'|^2$ is defined as

$$s_{cw}^{(i)}\ (K,N) = E\{(|s|^2)^i\}, \tag{3.12}$$

where $s_{cw}(K, N) = s_{cw}^{(1)}(K, N)$.

In the succeeding sections, expressions for $\sigma_{nw}^2(K,N)$, $\mathrm{SNR}_{nw}(K,N)$, $\sigma_{cw}^2(K,N)$, $s_{cw}(K,N)$, $\mathrm{NSR}_{cw}(K,N)$, $\mathrm{SNR}_{nw}^{(i)}\ (K,N)$, and $s_{cw}^{(i)}\ (K,N)$ are derived by using the GS canceller as an analysis tool.

IV. SMI AND GS CANCELLER EQUIVALENCE

In this section the SMI canceller and the GS canceller are shown to be equivalent in the sense that the SMI's estimated linear weighting vector is identical to the GS's equivalent estimated vector. Hence, if either concurrent or nonconcurrent processing is used, the output of the SMI and GS canceller is identical in the transient state (finite averaging). For this equivalence to be true, infinite computational accuracy and the nonsingularity of the estimated input covariance is assumed.

We briefly describe the SMI algorithm for the SLC configuration. If R_a is the $(N-1) \times (N-1)$ auxiliary input covariance matrix and $\mathbf{r}_{am}$ is the $N-1$ length cross-correlation vector of auxiliary channel against the main channel, the optimal (minimum output variance) $N-1$ length weighting vector $\mathbf{w}_a$ is given by

$$\mathbf{w}_a = R_a^{-1}\,\mathbf{r}_{am}\,. \tag{4.1}$$

For the SMI algorithm, R_a and $\mathbf{r}_{am}$ are estimated, the auxiliary weighting vector is calculated by using Eq. (4.1), this vector is applied to the auxiliary channels, and the resultant is subtracted from the main channel.

Define the input data vector

$$\mathbf{x}_a(k) = [x_1(k), x_2(k), \ldots, x_{N-1}(k)]^T, \quad k = 1, 2, \ldots, K \tag{4.2}$$

where a refers to the auxiliary channels and T denotes the vector (or matrix) transpose operation. The estimates of R_a and $\mathbf{r}_{am}$, denoted by $\hat{R}_a$ and $\hat{\mathbf{r}}_{am}$, are given by the expressions

$$\hat{R}_a = \frac{1}{K} \sum_{k=1}^{K} \mathbf{x}_a^*(k)\, \mathbf{x}_a^T(k) \tag{4.3}$$

and

$$\hat{\mathbf{r}}_{am} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{x}_a^*(k)\, x_0(k), \tag{4.4}$$

where the estimated linear weighting vector can be found by using the equation

$$\hat{R}_a \hat{\mathbf{w}} = \hat{\mathbf{r}}_{am}. \tag{4.5}$$

We define the following K-length input data vectors

$$\mathbf{x}_n = [x_n(1), x_n(2), \ldots, x_n(K)]^T, \quad n = 0, 1, \ldots, N-1 \tag{4.6}$$

and a $K \times (N-1)$ auxiliary input data matrix A, where

$$A = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N-1}). \tag{4.7}$$

It is straightforward to show that

$$\hat{R}_a = A^t A \tag{4.8}$$

and

$$\hat{\mathbf{r}}_{am} = A^t \mathbf{x}_0, \tag{4.9}$$

where t denotes the conjugate transpose. Thus, by using Eq. (4.5),

$$\hat{\mathbf{w}} = \hat{R}_a^{-1} \hat{\mathbf{r}}_{am} = (A^t A)^{-1} A^t \mathbf{x}_0. \tag{4.10}$$

Note that a necessary condition that $\hat{R}_a$ be nonsingular (and hence a unique solution for $\hat{\mathbf{w}}$ exists) is that $K \geq N-1$. To show this, assume that $K < N-1$ and define an $(N-1) \times (N-1)$ augmented matrix A_{aug} as

$$A_{aug} = \begin{bmatrix} A \\ ---- \\ \mathbf{0} \end{bmatrix}, \tag{4.11}$$

where the last $N - K - 1$ rows are zero filled. Now

$$\hat{R}_a = A^t_{aug} A_{aug}. \tag{4.12}$$

However the determinant of $\hat{R}_a$, denoted by $\det(\hat{R}_a)$, equals $\det(A_{aug}) \cdot \det(A^t_{aug})$. Since $\det(A_{aug}) = 0$, it follows that $\det(\hat{R}_a) = 0$ so that $\hat{R}_a$ is singular if $K < N - 1$.

We now show that if $K = N - 1$, that output noise residue is zero for the concurrent processor implementation. Set $\hat{\mathbf{w}} = (\hat{w}_1, \hat{w}_2, \ldots \hat{w}_{N-1})$ and let $\mathbf{z}$ be the K length output residue vector of the concurrent canceller. As a result

$$\mathbf{z} = \mathbf{x} - \sum_{N=1}^{N-1} \hat{w}_n \mathbf{x}_n. \tag{4.13}$$

By using Eq. (4.10), it can be shown that

$$\begin{aligned} \mathbf{z} &= \mathbf{x}_0 - A(A^t A)^{-1} A^t \mathbf{x}_0 \\ &= (I_K - A(A^t A)^{-1} A^t)\mathbf{x}_0, \end{aligned} \tag{4.14}$$

where I_K denotes the $K \times K$ identity matrix. For $K = N-1$, $A(A^t A)^{-1} A^t = I_K$, so that $\mathbf{z} = 0$.

As a result of the preceding discussion, in the following development for nonconcurrent processing we restrict $K \geq N - 1$, and for concurrent processing we restrict $K \geq N$.

If we set $\hat{\mathbf{w}} = (\hat{w}_1, \hat{w}_2, \ldots, \hat{w}_{N-1})^T$, then it can be shown that Eq. (4.5) reduces to solving the following system of linear equations:

$$\sum_{n=1}^{N-1} \mathbf{x}_1^t \mathbf{x}_n \hat{w}_n = -\mathbf{x}_1^t \mathbf{x}_0, \tag{4.15}$$

$$\sum_{n=1}^{N-1} \mathbf{x}_2^t \mathbf{x}_n \hat{w}_n = -\mathbf{x}_2^t x_0,$$

$$\vdots$$

$$\sum_{n=1}^{N-1} \mathbf{x}_{N-1}^t \mathbf{x}_n \hat{w}_n = -\mathbf{x}_{N-1}^t \mathbf{x}_0.$$

We show that the solution for the SMI weights, $\hat{w}_1, \hat{w}_2, \ldots, \hat{w}_{N-1}$, which orthogonalizes the auxiliary input data vectors $\mathbf{x}_n$, $n = 1, 2, \ldots, N-1$ to the output residue vector $\mathbf{z}$, is obtained by solving the system of equations given by Eq. (4.15); i.e., the condition

$$\mathbf{x}_n^t \, \mathbf{z} = 0, \qquad n = 1, 2, \ldots, N-1 \tag{4.16}$$

results in a system of equations identical to Eq. (4.15). Note that Eqs. (4.13) and (4.16) imply concurrent processing. However the weight calculation is valid for either concurrent or nonconcurrent processing.

If the concurrent GS canceller orthogonalizes the output data vector with respect to the auxiliary input data vectors $\mathbf{x}_n$, $n = 1, 2, \ldots, N-1$, it follows that the equivalent linear weighting vector associated with the GS structure is identical to that computed for the SMI algorithm. We show that a concurrent GS canceller orthogonalizes the output data vector $\mathbf{z}$ with respect to the auxiliary input data vector by using mathematical induction. This is obviously true for $N = 2$; we assume that it is true for all integers less than $N-1$ and demonstrate that it is true for when the number of channels equals N.

From Fig. 2 and our assumptions:

$$\mathbf{x}_n^{(2)t} \mathbf{z} = 0, \qquad n = 1, 2, \ldots, N-2 \tag{4.17}$$

and

$$\mathbf{x}_{N-1}^{(1)t} \, \mathbf{x}_n^{(2)} = 0, \qquad n = 0, 1, 2, \ldots, N-2, \tag{4.18}$$

where $\mathbf{x}_n^{(1)}$, $\mathbf{x}_n^{(2)}$ are the K-length data vectors associated with $x_n^{(1)}$, $x_n^{(2)}$, respectively. Furthermore, the output vector can be written as

$$\mathbf{z} = \mathbf{x}_0^{(2)} - \sum_{n=1}^{N-2} W_n' \, \mathbf{x}_n^{(2)}, \tag{4.19}$$

where $W_1', W_2', \ldots, W_{N-2}'$ is representative of the equivalent linear weighting of the input vectors from level 2 through $N-1$ of the GS structure (see Fig. 1(a)). Using Eq. (2.1) in Eq. (4.17) results in

$$\mathbf{x}_n^{(2)t} \mathbf{z} = (\mathbf{x}_n^{(1)} - w_n^{(1)} \mathbf{x}_{N-1}^{(1)})^t \, \mathbf{z} = \mathbf{x}_n^{(1)t} \mathbf{z} - w_n^{(1)*} \mathbf{x}_{N-1}^{(1)t} \, \mathbf{z} = 0, \quad n = 1, 2, \ldots, N-2. \tag{4.20}$$

From Eqs. (4.18) and (4.19) it can be shown that

$$\mathbf{x}_{N-1}^{(1)t} \mathbf{z} = \mathbf{x}_{N-1}^{(1)t} \, \mathbf{x}_0^{(2)} - \sum_{n=1}^{N-2} W_n' \mathbf{x}_n^{(2)t} \mathbf{x}_{N-1}^{(1)} = 0. \tag{4.21}$$

Thus, from Eqs. (4.20) and (4.21), it follows that

$$\mathbf{x}_n^{(1)t} \mathbf{z} = 0, \qquad n = 1, 2, \ldots, N-1. \tag{4.22}$$

Since $\mathbf{x}_n = \mathbf{x}_n^{(1)}$, $n = 1, 2, \ldots, N-1$, we have shown that the auxiliary input data vectors are orthogonal to the output vector.

V. INVARIANT TRANSFORMS AND GS THEOREMS

In this section, we discuss two types of matrix transforms on the input data that significantly simplify the forthcoming analysis. In addition, two theorems related to concurrent GS processing are presented. Let C be any $N \times N$ nonsingular matrix. It is well known [7] that transforming the input channels $x_M, x_1, \ldots, x_{N-1}$ by this transform does not change the transient or steady state performance of the SMI (or GS) canceller. GS cancellation is equivalent to a specific matrix transformation of the input channels. For the GS canceller, the transform matrix C has the upper triangular matrix form. Figure 4 shows an equivalent configuration of a GS canceller in the transient state. Here C is implemented by passing the input channels through a $GS_{\infty,N}$ structure followed by a power equalizer on the output auxiliary channels. The output powers of the AUX channels after power equalization are equal to σ^2_{min}. Note that each input channel into the $GS_{K,N}$ structure is orthogonal in the steady state to the other channels and that all input channels have the same power level, σ^2_{min}. Also, without loss of generality we can define $\sigma^2_{min} = 1$.

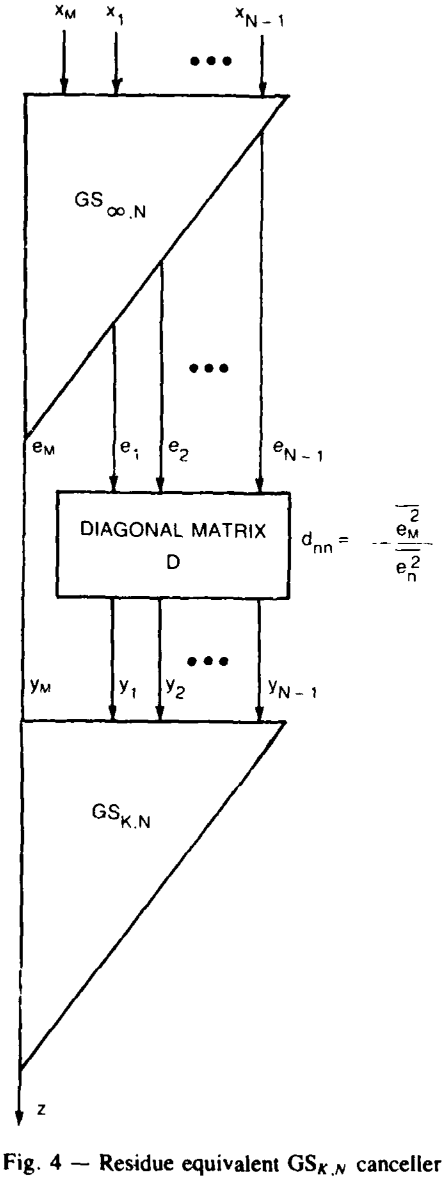

Fig. 4 — Residue equivalent $GS_{K,N}$ canceller using the power equalizer matrix

The structure shown in Fig. 4 illustrates that any GS canceller-structure can be divided into two parts—a deterministic steady-state, front-end processor whereby the main channel is decorrelated from the auxiliary channels and a stochastic back-end processor that is driven by uncorrelated equal-powered noise in each channel. The back-end processor is independent of the input covariance matrix, and the auxiliary weights associated with the back-end processor go to zero as $K \to \infty$. Hence the convergence properties of the GS canceller can be studied by analyzing the convergence properties of the back-end processor. Thus from this point on, the input channels are assumed to be orthogonal and of equal power.

A second matrix transform that significantly simplifies the forthcoming analysis is now discussed. Let Φ be any $K \times K$ unitary matrix, i.e., $\Phi^t \Phi = I_K$. Let us transform each input channel noise vector $\mathbf{x}_n$, $n = 0, 1, 2, \ldots, N-1$ by Φ such that

$$\mathbf{x}'_n = \Phi\, \mathbf{x}_n, \qquad n = 0, 1, \ldots, N-1, \tag{5.1}$$

where $\mathbf{x}'_n$, $n = 0, 1, \ldots, N-1$ is the resultant output noise vector. If we input this noise vector into a $GS_{K,N}$ canceller, we can show that the estimated weights using the $\mathbf{x}_n$ inputs are identical to those using the $\mathbf{x}'_n$ inputs. This is easily proved by substituting $\mathbf{x}'_n$ as given by Eq. (5.1) into Eq. (4.7), which is the system of equations that solves for the auxiliary weights. Because

$$\mathbf{x}'^t_n\, \mathbf{x}'_m = (\Phi \mathbf{x}_n)^t\, (\Phi \mathbf{x}_m) = \mathbf{x}^t_n\, \mathbf{x}_m, \qquad \text{for any } n, m, \tag{5.2}$$

an identical system of linear equations results. Thus, the estimated weights are identical.

One simplification that readily presents itself because of the above invariant transform pertains to the signal representation for concurrent processing. Let the input signal be represented by the K-length vector s, where

$$\mathbf{s} = (s_1, s_2, \ldots, s_K)^T. \tag{5.3}$$

It is known that a unitary matrix transform Φ_s exists that transforms s into a K-length vector with a nonzero first element and all other elements equal to zero. In fact

$$\tilde{\mathbf{s}} = \Phi_s \mathbf{s} = (\sqrt{\mathbf{s}^t \mathbf{s}}, 0, 0, \ldots, 0)^T. \tag{5.4}$$

Thus we need only consider an input signal vector that is proportional to the form $(1, 0, \ldots, 0)^T$ and $\sqrt{\mathbf{s}^t \mathbf{s}}$.

The following two theorems illustrate the proficiency of the concurrent GS canceller in performing the "cancellation operation."

Theorem 1: If

$$x_M = \sum_{n=1}^{N-1} c_n x_n, \tag{5.5}$$

then the $GS_{K,N}$ structure using concurrent processing cancels x_M exactly regardless of the number of input samples K per channel.

Proof: We use proof by induction. First we prove the theorem is true for $N = 2$; then we show that if it is true for $N - 1$, it is also true for N.

For $N = 2$, $x_M = c_1 x_1$, let w be the decorrelation weight derived from K input samples in each channel, i.e.,

$$w = \frac{\sum_{k=1}^{K} x_1^*(k) x_0(k)}{\sum_{k=1}^{K} |x_1(k)|^2}. \tag{5.6}$$

Substituting $x_M = c_1 x_1$ into Eq. (5.6) gives $w = c_1$, and therefore $x_M - w x_1 = 0$. Thus, Theorem 1 is true for $N = 2$.

Assume that the theorem is true for all integers less than N. Now $x_n^{(2)}$, $n = 0, 1, \ldots, N - 2$ are the outputs from the first level of $GS_{K,N}$ as illustrated in Fig. 1(a). That is

$$x_n^{(2)} = x_n^{(1)} - w_n^{(1)} x_{N-1}^{(1)}, \quad n = 0, 1, \ldots, N - 2 \tag{5.7}$$

where

$$w_n^{(1)} = \frac{\sum_{k=1}^{K} x_{N-1}^{(1)*}(k) x_n^{(1)}(k)}{\sum_{k=1}^{K} |x_{N-1}^{(1)}(k)|^2}. \tag{5.8}$$

For $w_0^{(1)}$, substituting Eq. (5.5) into Eq. (5.8) gives

$$w_0^{(1)} = \sum_{n=1}^{N-2} c_n w_n^{(1)} + c_{N-1}. \tag{5.9}$$

Hence,

$$\begin{aligned} x_0^{(2)} = x_0^{(1)} - w_0^{(1)} x_{N-1}^{(1)} &= \sum_{n=1}^{N-2} c_n x_n^{(1)} + c_{N-1} x_{N-1}^{(1)} - \left(\sum_{n=1}^{N-2} c_n w_n + c_{N-1} \right) x_{N-1}^{(1)}, \\ &= \sum_{n=1}^{N-2} c_n \left(x_n^{(1)} - w_n x_{N-1}^{(1)} \right), \\ &= \sum_{n=1}^{N-2} c_n x_n^{(2)}. \end{aligned} \tag{5.10}$$

Figure 5 illustrates the form of the output after first-level processing. Note that $x_0^{(2)}, x_1^{(2)}, \ldots, x_{N-2}^{(2)}$ are the inputs to a $GS_{K,N-1}$ structure. Because $x_0^{(2)}$ is a linear sum of $x_1^{(2)}, x_2^{(2)}, \ldots, x_{N-2}^{(2)}$ as shown by Eq. (5.10), and the theorem holds for $N - 1$, then $x_0^{(2)}$ is exactly cancelled independent of K. Hence the theorem holds for N and the theorem is proved by induction.

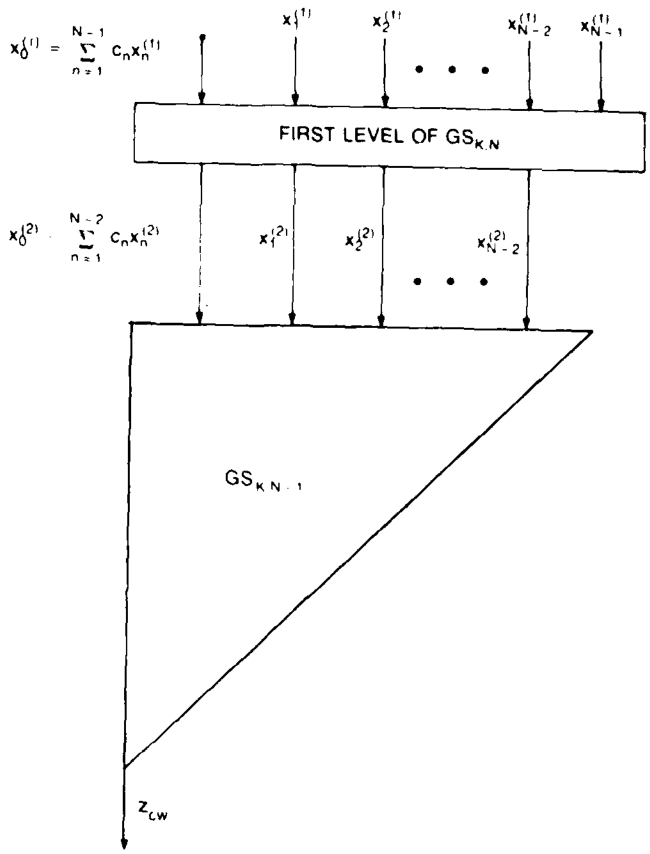

Fig. 5 — Concurrent GS canceller after first-level processing

A theorem that follows as a result of Theorem 1 is

Theorem 2: If

$$x_M = \sum_{n=1}^{N-1} c_n x_n + e , \tag{5.11}$$

then the output noise z_{cw} $GS_{K,N}$ *structure using concurrent processing depends only on* e, *i.e.,* z_{cw} *is independent of* $\sum_{n=1}^{N-1} c_n x_n$.

Proof: Let us write

$$x_M = x_{01} + x_{02} , \tag{5.12}$$

where

$$x_{01} = \sum_{n=1}^{N-1} c_n x_n ; \quad x_{02} = e . \tag{5.13}$$

From Fig. 1(a), all the $w_n^{(m)}$, $n \geq 1$ are independent of x_M, x_{01}, or x_{02}, i.e., all computed weights to the right of the main channel do not depend on the main channel inputs.

For $w_0^{(1)}$ (the first level, main channel weight),

$$w_0^{(1)} = \frac{\sum_{k=1}^{K} x_{N-1}^{(1)*}(k)\,[x_{01}^{(1)}(k) + x_{02}^{(1)}(k)]}{\sum_{k=1}^{K} |x_{N-1}^{(1)}(k)|^2}$$

$$= w_{01}^{(1)} + w_{02}^{(1)}\,, \tag{5.14}$$

where

$$w_{01}^{(1)} = \frac{\sum_{k=1}^{K} x_{N-1}^{(1)*}(k)\, x_{01}^{(1)}(k)}{\sum_{k=1}^{K} |x_{N-1}^{(1)}(k)|^2}\,; \tag{5.15a}$$

$$w_{02}^{(1)} = \frac{\sum_{k=1}^{K} x_{N-1}^{(1)*}(k)\, x_{02}^{(1)}(k)}{\sum_{k=1}^{K} |x_{N-1}^{(1)}(k)|^2}\,. \tag{5.15b}$$

Thus

$$x_0^{(2)} = x_{01}^{(1)} + x_{02}^{(1)} - (w_{01}^{(1)} + w_{02}^{(1)})\, x_{N-1}^{(1)}\,,$$

$$= [x_{01}^{(1)} - w_{01}^{(1)}\, x_{N-1}^{(1)}] + [x_{02}^{(1)} - w_{02}^{(1)}\, x_{N-1}^{(1)}]\,,$$

$$= x_{01}^{(2)} + x_{02}^{(2)}\,, \tag{5.16}$$

where

$$x_{01}^{(2)} = x_{01}^{(1)} - w_{01}^{(1)}\, x_{N-1}^{(1)}\,;\quad x_{02}^{(2)} = x_{02}^{(1)} - w_{02}^{(1)} x_{N-1}^{(1)}\,. \tag{5.17}$$

Note that the outputs $x_{01}^{(2)}$ and $x_{02}^{(2)}$ could have been computed independently of each other. Each output depends only on its respective input, x_{01} or x_{02}. We can continue down the main channel, computing the outputs of each successive two-input GS as the sum of the residues that result from components of the input sum.

As a result, an equivalent GS canceller for a main channel that is the sum of two components is shown in Fig. 6. Here each individual $GS_{K,N}$ structure is computed with respect to its main channel input with the outputs at each $GS_{K,N}$ summed to form the resultant. Hence, because of Theorem 1, x_{01} is nulled completely so that the resultant depends only on x_{02} or e. Thus, we need consider only the independent noise term of x_M when computing the output noise power.

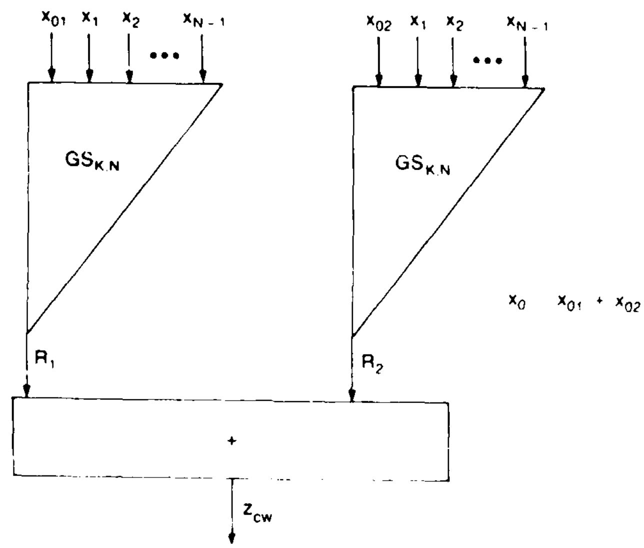

Fig. 6 — Equivalent concurrent GS canceller for main channel that is the sum of two components

VI. THE TWO-INPUT GS CANCELLER

The basis for understanding the convergence properties of a GS canceller begins with studying the two-input GS canceller illustrated in Fig. 7 where from the discussion in the previous section, x_0 and x_1 are r.v.'s that are assumed to be equal powered and uncorrelated. Let

$$\mathbf{z}_{cw} = (z_{cw}(1), z_{cw}(2), \ldots, z_{cw}(K))^T$$

be the concurrent output noise residue vector and Z_{nw} be the nonconcurrent output noise residue. For a two-input GS canceller

$$\mathbf{z}_{cw} = \mathbf{x}_0 - \hat{w}\,\mathbf{x}_1, \tag{6.1}$$

$$Z_{nw} = X_0 - \hat{w}\,X_1, \tag{6.2}$$

where

$$\hat{w} = \frac{\mathbf{x}_1^t \mathbf{x}_0}{\mathbf{x}_1^t \mathbf{x}_1}. \tag{6.3}$$

Furthermore, the transient output noise powers are given by

$$\hat{\sigma}_{cw}^2 = \frac{1}{K}\,\frac{\mathbf{z}_{cw}^t \mathbf{z}_{cw}}{\sigma_{\min}^2} = \frac{1}{K\sigma_{\min}^2}\left(\mathbf{x}_0^t\mathbf{x}_0 - \frac{|\mathbf{x}_1^t\,\mathbf{x}_0|^2}{\mathbf{x}_1^t\mathbf{x}_1}\right) \tag{6.4}$$

and

$$\overline{\hat{\sigma}_{nw}^2} = 1 + |\hat{w}|^2 = 1 + \frac{|\mathbf{x}_1^t\mathbf{x}_0|^2}{(\mathbf{x}_1^t\mathbf{x}_1)^2}. \tag{6.5}$$

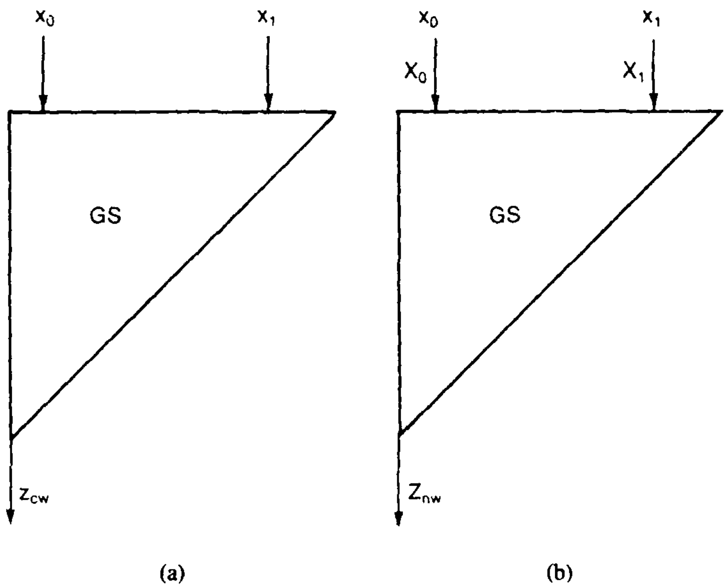

Fig. 7 — Two-input GS cancellers

In Appendix A (and Ref. 8), it is shown under Assumptions 1 to 6 given in Section II that $\overline{\hat{\sigma}_{nw}^2} = \eta$ has the following probability density function.

$$p(\eta) = \frac{K}{\eta^{K+1}}, \qquad \eta \geq 1. \tag{6.6}$$

Now the transient SNR is equal to the reciprocal of $\overline{\hat{\sigma}_{nw}^2}$. Thus if SNR $= 1/\overline{\hat{\sigma}_{nw}^2} = \rho$, then by the use of elementary probability theory,

$$p(\rho) = K\rho^{K-1}, \qquad 0 \leq \rho \leq 1. \tag{6.7}$$

By using Eqs. (6.6) and (6.7), it can be shown that

$$\sigma_{nw}^2\ (K,2) = 1 + \frac{1}{K-1}, \tag{6.8}$$

$$\mathrm{SNR}_{nw}\ (K,2) = \frac{K}{K+1}, \tag{6.9}$$

and

$$\mathrm{SNR}_{nw}^{(i)}(K,2) = \frac{K}{K+i}. \tag{6.10}$$

From Eq. (6.4), the expected value of $\hat{\sigma}_{cw}^2$ conditioned on $\mathbf{x}_1$ is given by

$$E\{\hat{\sigma}_{cw}^2 \mid \mathbf{x}_1\} = \frac{1}{K\sigma_{\min}^2}\left[E\{\mathbf{x}_0^t\mathbf{x}_0\} - \frac{\mathbf{x}_1^t E\{\mathbf{x}_0\mathbf{x}_0^t\}\mathbf{x}_1}{\mathbf{x}_1^t\mathbf{x}_1}\right]. \tag{6.11}$$

By assumption $E\{\mathbf{x}_0\mathbf{x}_0^t\} = \sigma_{min}^2 I_K$, where I_K is the $K \times K$ identity matrix and $E\{\mathbf{x}_0^t\mathbf{x}_0\} = K\sigma_{min}^2$. Thus Eq. (6.11) reduces to

$$E\{\hat{\sigma}_{cw}^2 \mid \mathbf{x}_1\} = 1 - \frac{1}{K}. \tag{6.12}$$

Note that this expression is independent of $\mathbf{x}_1$ so that

$$E\{\hat{\sigma}_{cw}^2\} = 1 - \frac{1}{K} \tag{6.13}$$

or

$$\sigma_{cw}^2(K,2) = 1 - \frac{1}{K}. \tag{6.14}$$

Also note the assumption that the inputs are Gaussian r.v.'s was not used in this derivation. In fact each data point in either channel can have any p.d.f. so long as it has a zero mean and identical variances.

The unnormalized output signal vector through the two-input GS canceller is given by the expression

$$\mathbf{s}'' = \left\{ I_K - \frac{\mathbf{x}_1\mathbf{x}_1^t}{\mathbf{x}_1^t\mathbf{x}_1} \right\} \mathbf{s}. \tag{6.15}$$

Hence the sample average output signal power is given by

$$|\mathbf{s}''|^2 = \frac{1}{K}\mathbf{s}''^t\mathbf{s}'' = \frac{1}{K}\left\{ \mathbf{s}^t\mathbf{s} - \frac{|\mathbf{s}^2\mathbf{x}_1|^2}{\mathbf{x}_1^t\mathbf{x}_1} \right\}. \tag{6.16}$$

We show in Section V that s and $\mathbf{x}_1$ can be transformed by a unitary matrix without affecting the resultant output measures. If we transform the s and $\mathbf{x}_1$ by the unitary Φ_s matrix defined in Section V, then Eq. (6.16) reduces to

$$|s''|^2 = \frac{1}{K}\mathbf{s}^t\mathbf{s}\left[1 - \frac{|x_1(1)|^2}{\sum\limits_{k=1}^{K} |x_1(k)|^2} \right]. \tag{6.17}$$

Now the normalized signal power is given by

$$|s'|^2 = \frac{|s''|^2}{\frac{1}{K}\mathbf{s}^t\mathbf{s}}, \tag{6.18}$$

so that

$$|s'|^2 = 1 - \frac{|x_1(1)|^2}{\sum_{k=1}^{K} |x_1(k)|^2}. \tag{6.19}$$

It is straightforward to show that if x_1 is a normalized K-length, multivariate, complex circular Gaussian vector, then $\xi = |s'|^2$ has the following p.d.f.

$$p(\xi) = (K - 1)\xi^{K-2}, \quad \xi \geq 0. \tag{6.20}$$

Furthermore, the moments of $|s'|^2$ can be found by using the above p.d.f. and are given by

$$s_{cw}^{(i)}(K, 2) = \frac{K-1}{K-1+i}. \tag{6.21}$$

From Eq. (6.21) it follows that

$$s_{cw}(K,2) = 1 - \frac{1}{K}.$$

Finally, we can show that

$$\mathrm{N\hat{S}R}_{cw} = \frac{\hat{\sigma}_{cw}^2}{|s'|^2} = \frac{\frac{1}{K}(\mathbf{x}_0^t\mathbf{x}_0 - |\mathbf{x}_1^t\mathbf{x}_0|^2/\mathbf{x}_1^t\mathbf{x}_1)}{1 - \frac{|x_1(1)|^2}{\sum_{k=1}^{K}|x_1(k)|^2}}. \tag{6.22}$$

If $\mathrm{N\hat{S}R}_{cw}$ is averaged over $\mathbf{x}_0$, then

$$E\{\mathrm{N\hat{S}R}_{cw} \mid \mathbf{x}_1\} = \frac{K-1}{K} \cdot \frac{1}{1 - \frac{|x_1(1)|^2}{\sum_{k=1}^{K}|x_1(k)|^2}}$$

$$= \frac{K-1}{K}\left[1 + \frac{|x_1(1)|^2}{\sum_{k=2}^{K}|x_1(k)|^2}\right]. \tag{6.23}$$

Now since $\mathbf{x}_1$ is a normalized K-length, multivariate, complex circular Gaussian vector,

$$E\{|x_1(1)|^2\} = 1 \quad \text{and} \quad E\left\{\frac{1}{\sum_{K=2}^{K}|x_1(k)|^2}\right\} = \frac{1}{K-2}. \tag{6.24}$$

Thus

$$\mathrm{NSR}_{cw}\ (K,2) = E\{\mathrm{N\hat{S}R}_{cw}\} = \frac{K-1}{K}\left(1+\frac{1}{K-2}\right)$$

$$= \frac{(K-1)^2}{K(K-2)}. \tag{6.25}$$

VII. GENERAL MOMENT THEOREM FOR GS CANCELLERS

Let $\hat{A}_{K,N}$ denote a transient unnormalized moment of an output measure (output noise power residue or SNR) associated with a concurrent or nonconcurrent GS canceller with N input channels and K independent samples per channel. Define the normalized average transient moment as

$$A(K,N) = \frac{E\{\hat{A}_{K,N}\}}{\lim\limits_{K\to\infty} E\{\hat{A}_{K,N}\}}, \tag{7.1}$$

where $A(K,1$. In this section we prove the following theorem:

General Moment Theorem for GS Cancellers: If assumptions 1 to 6 hold, then

$$A(K,N) = A(K,2)\,A(K-1,N-1) \tag{7.2}$$

or equivalently

$$A(K,N) = \prod_{k=K-N+2}^{K} A(k,2). \tag{7.3}$$

Proof. We prove this by mathematical induction. First, the theorem is obviously true for $N = 2$. Thus, we can assume that the theorem is true for all integers less than or equal to some upper bound: $N-1$. We can then show that it is true for N, which implies that it is true for any $N \geq 2$.

Again assume that all input channels are of equal power and uncorrelated. It is shown in Ref. 7 and discussed in Section V that this assumption does not change the output measures. A $\mathrm{GS}_{K,N}$ structure can be decomposed as shown in Fig. 8 into a first-level processor followed by a $\mathrm{GS}_{K,N-1}$ structure. The output data K-length vectors of the first-level processor can be written as

$$\mathbf{y}_n = \mathbf{x}_n - \hat{w}_n \mathbf{x}_{N-1}, \qquad \hat{w} = \frac{\mathbf{x}_{N-1}^t \mathbf{x}_n}{\mathbf{x}_{N-1}^t \mathbf{x}_{N-1}}, \qquad n = 0, 1, \ldots, N-2 \tag{7.4}$$

or

$$\mathbf{y}_n = \mathbf{x}_n - \frac{\mathbf{x}_{N-1}^t \mathbf{x}_n}{\mathbf{x}_{N-1}^t \mathbf{x}_{N-1}}\, \mathbf{x}_{N-1}$$

or

$$\mathbf{y}_n = \left(I_K - \frac{\mathbf{x}_{N-1}\mathbf{x}_{N-1}^t}{\mathbf{x}_{N-1}^t \mathbf{x}_{N-1}}\right)\mathbf{x}_n, \qquad n = 0, 1, 2, \ldots, N-1. \tag{7.5}$$

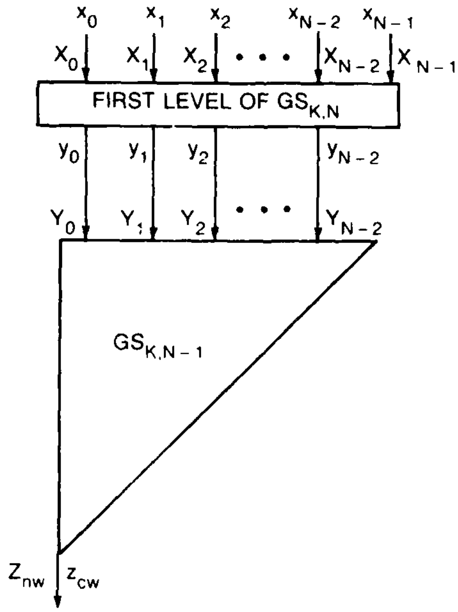

Fig. 8 — Decomposition of $GS_{K,N}$

It can be shown that

$$I_K - \frac{\mathbf{x}_{N-1}\mathbf{x}_{N-1}^t}{\mathbf{x}_{N-1}^t\mathbf{x}_{N-1}} = \Phi^t \Lambda \Phi , \tag{7.6}$$

where Φ is a $K \times K$ unitary matrix and Λ is a diagonal matrix where the first element is 0 and all other diagonal elements are equal to 1. Thus

$$\mathbf{y}_n = \Phi^t \ \Lambda \ \Phi \mathbf{x}_n , \qquad n = 0, 1, \ldots, N - 2. \tag{7.7}$$

As shown in Section V, the output data set $\mathbf{y}_n$, $n = 0, 1, \ldots, N - 2$ can be transformed by a unitary matrix Φ and not change the equivalent transient weighting vector of the $GS_{K,N-1}$ structure. Thus

$$\mathbf{u}_n = \Phi \mathbf{y}_n = \Lambda \ \Phi \mathbf{x}_n , \qquad n = 0, 1, \ldots, N - 2. \tag{7.8}$$

Now set $\mathbf{v}_n = \Phi \mathbf{x}_n$. Because $\mathbf{x}_n$ is a normalized K-length multivariate complex circular Gaussian vector, then $\mathbf{v}$ is the same. As a result, using the form of Λ and setting $\mathbf{u}_n = (u_{n1}, u_{n2}, \ldots, u_{nk})^T$, $\mathbf{v}_n = (v_{n1}, v_{n2}, \ldots, v_{nk})^T$, it follows from Eq. (7.8) that

$$u_{n1} = 0 \tag{7.9}$$

and

$$u_{nk} = v_{nk} , \qquad k = 2, 3, \ldots, K .$$

Hence, the input r.v.'s to the $GS_{K,N-1}$ structure are identically distributed to the input r.v.'s to the $GS_{K,N}$ structure except that their number has been reduced by one as illustrated in Fig. 9.

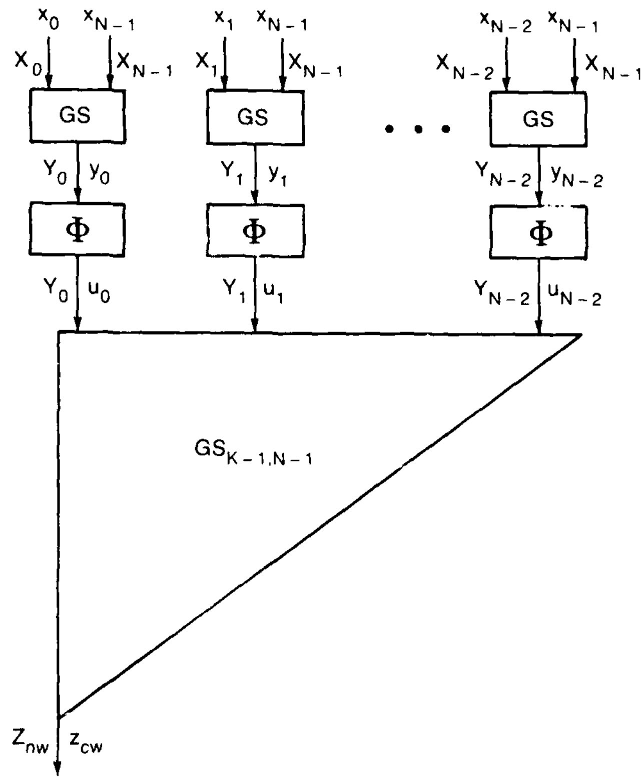

Fig. 9 — Further decomposition of $GS_{K,N}$

Consider the implications of this new structure by using concurrent processing. Let $\hat{B}_{K-1,N-1}$ denote a transient unnormalized moment of any output measure (output noise residue or SNR) associated with the $GS_{K-1,N-1}$ structure. Note that $\hat{A}_{K,N} = \hat{B}_{K-1,N-1}$. Then according to the General Moment Theorem,

$$\frac{E\{\hat{B}_{K-1,N-1} \mid \mathbf{x}_{N-1}\}}{\lim\limits_{K\to\infty} E\{\hat{B}_{K-1,N-1} \mid \mathbf{x}_{N-1}\}} = A(K-1, N-1). \tag{7.10}$$

Note in the limit taken above that K goes to infinity only in the $GS_{K-1,N-1}$ structure and not in the first-level processor that precedes the $GS_{K-1,N-1}$ structure. Thus

$$E\{\hat{A}_{K,N} \mid \mathbf{x}_{N-1}\} = A(K-1, N-1) \lim_{K\to\infty} E\{\hat{B}_{K-1,N-1} \mid \mathbf{x}_{N-1}\} \tag{7.11}$$

or

$$E\{\hat{A}_{k,N}\} = A(K-1, N-1) \lim_{K\to\infty} E\{\hat{B}_{K-1,N-1}\}. \tag{7.12}$$

The limit in Eq. (7.12) can be evaluated simply. It is equal to the associated moment of the output measure coming out of the first-level processor in the main channel or

$$\lim_{K\to\infty} \{\hat{B}_{K-1,N-1}\} = A(K,2) \lim_{K\to\infty} E\{\hat{A}_{K,N}\}. \tag{7.13}$$

Now, the limit on the right-hand side of Eq. (7.13) is equal to $\sigma^2_{\min} = 1$. Hence the theorem is proved for concurrent processing.

We now prove the theorem for nonconcurrent processing. The nonconcurrent data sample after first-level processing can be written as

$$Y_n = X_n - \hat{w}_n X_{N-1}, \quad n = 0, 1, \ldots, N-1, \tag{7.14}$$

where $\hat{w}_n$ is given in Eq. (7.4). Given $\mathbf{x}_{N-1}$ and X_{N-1}, it can be shown that the Y_n, $n = 0, 1, \ldots, N-2$ are uncorrelated complex Gaussian random variables, i.e., $E\{Y_{n_1} Y^*_{n_2}\} = 0$ and $E\{|Y_n|^2\}$ equals a constant. Thus the Y_n, $n = 0, 1, \ldots, N-2$ conditioned on $\mathbf{x}_{N-1}$ and X_{N-1} are identically distributed as the X_n, $n = 0, 1, \ldots, N-1$. Also note that the Φ transformation does not change the equivalent linear weighting of the $\mathrm{GS}_{K,N-1}$ processor or its equivalent $\mathrm{GS}_{K-1,N-1}$ processor. Hence, the inputs (concurrent and nonconcurrent) into the $\mathrm{GS}_{K-1,N-1}$ processor are identically distributed as those into the $\mathrm{GS}_{K,N}$ processor.

Defining $\hat{B}_{K-1,N-1}$ as before, it follows from the General Moment Theorem that

$$\frac{E\{\hat{B}_{K-1,N-1} \mid \mathbf{x}_{N-1}, X_{N-1}\}}{\lim\limits_{K\to\infty} E\{\hat{B}_{K-1,N-1} \mid \mathbf{x}_{N-1}, X_{N-1}\}} = A(K-1, N-1). \tag{7.15}$$

By using Eq. (7.15) and reasoning similar to that used for proving the theorem for concurrent processing, the theorem follows for nonconcurrent processing.

VIII. CONVERGENCE RESULTS

In this section, the General Moment Theorem for GS Cancellers (Eq. (7.3)) is used to derive a number of results. Most of these results are demonstrated in Refs. 7 and 8.

Employing the General Moment Theorem under Assumptions 1 to 6 and the expressions for $\sigma^2_{cw}(K,2)$, $s_{cw}(K,2)$, $s^{(i)}_{cw}(K,2)$, $\mathrm{NSR}_{cw}(K,2)$, $\sigma^2_{nw}(K,2)$, $\mathrm{SNR}_{nw}(K,2)$, and $\mathrm{SNR}^{(i)}_{nw}(K,2)$ given in Section IV, we can show that

$$\sigma^2_{cw}(K,N) = 1 - \frac{N-1}{K}, \tag{8.1}$$

$$s_{cw}(K,N) = 1 - \frac{N-1}{K}, \tag{8.2}$$

$$s^{(i)}_{cw}(K,N) = \frac{(K-1)!\,(K+i-N)!}{(K-N)!\,(K+i-1)!}, \tag{8.3}$$

$$\mathrm{NSR}_{cw}(K,N) = \frac{(K-1)\,(K-N+1)}{K\,(K-N)}, \tag{8.4}$$

$$\sigma^2_{nw}(K,N) = \frac{K}{K-N+1}, \tag{8.5}$$

$$\mathrm{SNR}_{nw}(K,N) = \frac{K - N + 2}{K + 1}, \tag{8.6}$$

$$\mathrm{SNR}_{nw}^{(i)}(K,N) = \frac{K!\,(K + i - N + 1)!}{(K - N + 1)!\,(K + i)!}. \tag{8.7}$$

Equations (8.1) to (8.3) and (8.5) to (8.7) are given in Refs. 7 and 8.

Equation (8.7) can be used in ad hoc fashion to find the p.d.f of $\hat{\mathrm{SNR}}_{nw}$ for any K and N. If $\rho = \hat{\mathrm{SNR}}_{nw}$, then the p.d.f. that yields moments as given by Eq. (8.7) is

$$p(\rho) = \frac{K!}{(N - 2)!\,(K - N + 1)!}(1 - \rho)^{N-2}\,\rho^{K-N+1}, \quad 0 \le \rho \le 1. \tag{8.8}$$

From Eq. (8.8), the p.d.f. of $\overline{\hat{\sigma}_{nw}^2}$ can be obtained. Let $\eta = \overline{\hat{\sigma}_{nw}^2} = 1/\rho$. It is straightforward to show that

$$p(\eta) = \frac{K!}{(N - 2)!\,(K - N + 1)!}\,\frac{(\eta - 1)^{N-2}}{\eta^{K+1}}, \quad 1 < \eta \le \infty. \tag{8.9}$$

Note that Eqs. (8.8) and (8.9) were first derived in Refs. 7 and 8, respectively.

IX. THE CONCURRENT GS CANCELLER

The discussion and results in this section pertain specifically to the concurrent GS canceller.

A. Input-Output Matrix Transform

Consider the GS canceller with all of its output channels as shown in Fig. 2(b). The K-length input channel vectors are transformed into K-length orthogonal output vectors that form an orthogonal basis for the $N - 1$ auxiliary input vectors, i.e., if $\mathbf{z}_n = (z_n(1), z_n(2), \ldots, z_n(K))^T$, $n = 0, 1, \ldots, N - 1$ are the output channel vectors, then

$$\mathbf{z}_{n_1}^t \mathbf{z}_{n_2} = 0, \qquad n_1 \ne n_2. \tag{9.1}$$

It is elementary to show that the main channel output vector at the $n + 1$th level, $\mathbf{x}_0^{(n+1)}$, is given by the expression

$$\mathbf{x}_0^{(n+1)} = \left(I_K - \frac{\mathbf{z}_{N-n}\mathbf{z}_{N-n}^t}{\mathbf{z}_{N-n}^t\mathbf{z}_{N-n}} \right) \mathbf{x}_0^{(n)}. \tag{9.2}$$

Thus because $\mathbf{z}_{cw} = \mathbf{z}_0 = \mathbf{x}_0^{(N)}$ and $\mathbf{x}_0 = \mathbf{x}_0^{(1)}$, it follows that

$$\mathbf{z}_{cw} = \left[\prod_{n=1}^{N-1} \left(I_K - \frac{\mathbf{z}_n\mathbf{z}_n^t}{\mathbf{z}_n^t\mathbf{z}_n} \right) \right] \mathbf{x}_0. \tag{9.3}$$

Equation (9.3) can be simplified by using Eq. (9.1) as

$$\mathbf{z}_{cw} = \left[I_K - \sum_{n=1}^{N-1} \frac{\mathbf{z}_n \mathbf{z}_n^t}{\mathbf{z}_n^t \mathbf{z}_n} \right] \mathbf{x}_0 . \tag{9.4}$$

Set

$$G = I_K - \sum_{n=1}^{N-1} \frac{\mathbf{z}_n \mathbf{z}_n^t}{\mathbf{z}_n^t \mathbf{z}_n} . \tag{9.5}$$

The $K \times K$ matrix G is the input-output matrix transform. In addition, $G\mathbf{x}_n = \mathbf{0}$, $n = 1, 2, \ldots, N-1$. Thus, the rows of G have an orthonormal basis that is orthogonal to the orthonormal basis of $\mathbf{x}_n$, $n = 1, 2, \ldots, N-1$. If we write the auxiliary input samples as a $K \times (N-1)$ matrix, X_{aux}, then the orthonormal basis of the rows of G is the same as the orthonormal basis for the null space of C. A matrix having the form given by Eq. (9.5) is sometimes called a complementary projection matrix. Furthermore $G = G^t$, G is idempotent so that

$$G^2 = G . \tag{9.6}$$

B. Sample Average Bound

We use Eq. (9.4) to prove the following:

Output Power Sample Average Theorem: For concurrent GS cancellers, the sample average of the output power residue is always less than or equal to the sample average of the input power.

Proof: The input and output power sample averages are given by

$$\hat{\sigma}_{in}^2 = \frac{1}{K} \mathbf{x}_0^t \mathbf{x}_0 \tag{9.7}$$

$$\hat{\sigma}_{cw}^2 = \frac{1}{K} \mathbf{z}_{cw}^t \mathbf{z}_{cw} . \tag{9.8}$$

By use of Eqs. (9.4) and (9.6),

$$\begin{aligned} \hat{\sigma}_{cw}^2 &= \frac{1}{K} (G\mathbf{x}_0)^t G\mathbf{x}_0 \\ &= \frac{1}{K} \mathbf{x}_0^t G \mathbf{x}_0 \\ &= \frac{1}{K} \mathbf{x}_0^t \mathbf{x}_0 - \frac{1}{K} \sum_{n=1}^{N-1} \frac{|\mathbf{z}_n^t \mathbf{x}_0|^2}{\mathbf{z}_n^t \mathbf{z}_n} \\ &= \hat{\sigma}_{in}^2 - \frac{1}{K} \sum_{n=1}^{N-1} \frac{|\mathbf{z}_n^t \mathbf{x}_0|^2}{\mathbf{z}_n^t \mathbf{z}_n} . \end{aligned} \tag{9.9}$$

Since the summation term in Eq. (9.9) is always positive, the theorem follows.

C. P.D.F. of the Output Noise

Define Λ_n to be a $K \times K$ diagonal matrix with 0's in the first n diagonal elements and 1's in the rest. It is straightforward to show that G can be written in the form:

$$G = \Phi^t \Lambda_{N-1} \Phi, \tag{9.10}$$

where Φ is a $K \times K$ unitary matrix. Thus

$$\hat{\sigma}_{cw}^2 = \frac{1}{K} \mathbf{x}_0^t \Phi^t \Lambda_{N-1} \Phi \mathbf{x}_0 \tag{9.11}$$

$$= \frac{1}{K} (\Phi \mathbf{x}_0)^t \Lambda_{N-1} (\Phi \mathbf{x}_0).$$

We set a K-length vector, $\mathbf{u} = \Phi \mathbf{x}_0$. If $\mathbf{x}_0$ is a normalized K-multivariate complex Gaussian circular process, it can be shown that $\mathbf{u}$ is the same. Because of the form of Λ_{N-1},

$$\hat{\sigma}_{cw}^2 = \frac{1}{K} \sum_{k=N}^{K} |u_k|^2. \tag{9.12}$$

It is elementary to show that $K\hat{\sigma}_{cw}^2$ has a $K - N + 1$ order chi-square p.d.f. Thus if $\eta = \hat{\sigma}_{cw}^2$, then

$$p(\eta) = \frac{K^{K-N+1}}{(K - N)!} \eta^{K-N} e^{-K\eta}, \qquad \eta \geq 0. \tag{9.13}$$

Note that the output noise power p.d.f. is independent of the auxiliary p.d.f.'s and their independence properties. Only the main channel must be Gaussian, and its samples must be independent. Equation (9.13) is the same p.d.f. that was derived by Brennan and Reed in Ref. 8 under the assumption that all input channels are Gaussian.

Hence the following theorem results:

Concurrent GS Canceller Convergence Theorem: If the main channel of an N-input GS canceller consists of a K-length multivariate complex circular Gaussian vector of samples that are also independent of the non-time-coincident samples of the auxiliary channels (Assumption 3, Section II) and the auxiliary channels have arbitrary p.d.f.s, autocorrelation functions, and cross-correlation functions with other auxiliary channels, then the normalized output noise power residue $\eta = \hat{\sigma}_{cw}^2$ has a p.d.f given by Eq. (9.13).

D. P.D.F. and Power of the Signal

The input signal vector $\mathbf{s}$ is also transformed by the input-output matrix transform in the GS structure. It is shown in Section VIII that the moments of the output signal power, $|s'|^2$, are given by Eq. (8.3). It can be shown ad hoc that if the inputs satisfy Assumptions 1 to 6, then the p.d.f. of $\rho = |s'|^2$ is given by

$$p(\rho) = \frac{(K - 1)!}{(K - N)!\,(N - 2)!}\,(1 - \rho)^{N-2}\,\rho^{K-N}, \quad 0 < \rho < 1. \tag{9.14}$$

Again this is the same result that was derived in Ref. 8 although the signal model was different.

The input signal vector s is transformed by the projection matrix G defined by Eq. (9.5) into the unnormalized signal vector s′, given by the expression

$$\mathbf{s}' = \left[I_K - \sum_{n=1}^{N-1} \frac{\mathbf{z}_n \mathbf{z}_n^t}{\mathbf{z}_n^t \mathbf{z}_n} \right] \mathbf{s}. \tag{9.15}$$

Again without loss of generality, we can set $\mathbf{s} = (\sigma_s, 0, 0, \ldots, 0)^T$, where $\sigma_s^2 = \mathbf{s}^t\mathbf{s}$ is the sum of the input signal power across K samples. Thus by using Eq. (9.15),

$$\mathbf{s}'^t\mathbf{s}' = \sigma_s^2\, \mathbf{1}_0^T \left[I_K - \sum_{n=1}^{N-1} \frac{\mathbf{z}_n \mathbf{z}_n^t}{\mathbf{z}_n^t \mathbf{z}_n} \right] \mathbf{1}_0^T, \tag{9.16}$$

where $\mathbf{1}_0 = (1, 0, 0, \ldots, 0)^T$. Equation (9.16) simplifies to

$$\frac{\mathbf{s}'^t\mathbf{s}'}{\sigma_s^2} = 1 - \sum_{n=1}^{N-1} \frac{|z_n(1)|^2}{\sum_{k=1}^{K} |z_n(k)|^2}. \tag{9.17}$$

Now

$$s_{cw}(K,N) = E\left\{\frac{\mathbf{s}'^t\mathbf{s}'}{\sigma_s^2}\right\} = 1 - \sum_{n=1}^{N-1} E\left\{\frac{|z_n(1)|^2}{\sum_{k=1}^{K} |z_n(k)|^2}\right\}. \tag{9.18}$$

If the elements of $x_1, x_2, \ldots, x_{N-1}$ are identically distributed r.v.'s, then the elements of $z_1, z_2, \ldots, z_{N-1}$ are also identically distributed r.v.'s. It then follows that independent of the p.d.f.'s of these r.v.'s,

$$E\left\{\frac{|z_n(1)|^2}{\sum_{k=1}^{K} |z_n(k)|^2}\right\} = \frac{1}{K}. \tag{9.19}$$

Thus we have the theorem:

Concurrent GS Canceller Signal Theorem: If the samples of the auxiliary channels are identically distributed r.v.'s and the desired signal is only in the main channel, then

$$s_{cw}(K,N) = 1 - \frac{N-1}{K}. \tag{9.20}$$

E. General Result for Noise Power

We prove the following:

Concurrent Processing Theorem: Let an input data element in the vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{N-1}$ *have any p.d.f. (note each of the KN elements could have a different p.d.f.). Then under Assumptions 2 to 6, Section II,*

$$\sigma_{cw}^2 (K,N) = 1 - \frac{N-1}{K}. \tag{9.21}$$

Proof: Consider the form of $\hat{\sigma}_{cw}^2$ given by Eq. (9.9). We can show

$$E\{\hat{\sigma}_{cw}^2 \mid \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{N-1}\} = E\{\hat{\sigma}_{in}^2\} - \frac{1}{K} \sum_{n=1}^{N-1} \frac{\mathbf{z}_n^t E\{\mathbf{x}_0 \mathbf{x}_0^t\} \mathbf{z}_n}{\mathbf{z}_n^t \mathbf{z}_n}. \tag{9.22}$$

Now $E\{\hat{\sigma}_{in}^2\} = 1$ and $E\{\mathbf{x}_0 \mathbf{x}_0^t\} = I_K$. Hence

$$E\{\hat{\sigma}_{cw}^2 \mid \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{N-1}\} = 1 - \frac{N-1}{K}. \tag{9.23}$$

The theorem follows by integrating both sides of Eq. (9.23) over the joint p.d.f. of $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{N-1}$.

F. Noise-to-Signal Considerations

We show in Section VIII that the average NSR is given by the expression (repeated here)

$$\mathrm{NSR}(K,N) = \frac{(K-1)(K-N+1)}{K(K-N)}. \tag{9.24}$$

The number of samples $K_{3\mathrm{dB}}$, where the NSR is within 3 dB of the optimum, is found by setting NSR (K,N) equal to 2 and solving for $K_{3\mathrm{dB}}$. It is found that

$$K_{3\mathrm{dB}} = \frac{N + \sqrt{N^2 + 4(N-1)}}{2}. \tag{9.25}$$

However, note that for $K = N+1, N+2, N+3, N+4, N+5$,

$$\mathrm{NSR}(N+1, N) = 2 - \frac{2}{N+1},$$

$$\mathrm{NSR}(N+2, N) = 1.5 - \frac{1.5}{N+2},$$

$$\mathrm{NSR}(N+3, N) = 1.33 - \frac{1.33}{N+3}, \tag{9.26}$$

$$\text{NSR}\,(N+4, N) = 1.25 - \frac{1.25}{N+4},$$

$$\text{NSR}\,(N+5, N) = 1.2 - \frac{1.2}{N+5}.$$

Hence, for $K = N + 1$, the NSR is already within 3 dB of the optimum and for $K = N + 5$ is within 0.8 dB.

G. Discussion

The concurrent GS canceller converges rapidly as implied by Eq. (9.18). In fact it requires approximately N samples to achieve the 3 dB performance point, opposed to $2N$ samples required for the nonconcurrent processing [7]. However, other losses must be taken into account. These losses are best exemplified by considering the transformed output noise and output signal vectors after GS cancellation. These can be written as

$$\mathbf{s}' = (0, 0, \ldots, 0, s'_N, s'_{N+1}, \ldots, s'_K)^T, \tag{9.27}$$

$$\mathbf{n}' = (0, 0, \ldots, 0, n'_N, n'_{N+1}, \ldots, n'_K), \tag{9.28}$$

where the first $N - 1$ elements of each vector are equal to zero. Signal detection losses occur because we have lost $N - 1$ independent samples of signal and noise by going through the GS canceller. Hence, the signal detector after the GS canceller whether it be a coherent integrator or some other detector scheme (hypothesis test statistic generation and thresholding) has lost independent data samples, which decreases the probability of detection of the signal. Thus even though $K = N + 1$ samples yields an NSR per time sample that is within 3 dB of the optimal steady state ($K \to \infty$), the output samples residue are so correlated that they are equivalent to having only two independent samples out of the $N + 1$ output samples. As a result, signal detection after GS cancellation can degrade significantly.

X. LOWER BOUND

In this section, we derive a lower bound associated with convergence of a nonconcurrent GS canceller when the input data are not necessarily Gaussian. In the analysis, Assumptions 2 to 6 of *Section II hold. An element of the input data vectors* $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{N-1}$ can have any p.d.f.

A result for the two-input GS canceller is first established. For this case, it was previously shown that

$$\overline{\hat{\sigma}^2_{nw}} = 1 + |\hat{w}|^2 = 1 + \frac{\mathbf{x}_1^t \mathbf{x}_0 \mathbf{x}_0^t \mathbf{x}_1}{(\mathbf{x}_1^t \mathbf{x}_1)^2}. \tag{10.1}$$

We can show that because $E\{\mathbf{x}_0 \mathbf{x}_0^t\} = \sigma^2_{\min} I_K$,

$$E\{\overline{\hat{\sigma}^2_{nw}} \,|\, \mathbf{x}_1\} = 1 + \frac{\sigma^2_{\min}}{\mathbf{x}_1^t \mathbf{x}_1}. \tag{10.2}$$

Thus

$$\sigma_{nw}^2(K,2) = \overline{E\{\hat{\sigma}_{nw}^2\}} = 1 + \sigma_{\min}^2 \cdot E\left\{\frac{1}{\mathbf{x}_1^t\mathbf{x}_1}\right\}. \tag{10.3}$$

Appendix B shows that if z is any r.v. with a nonzero mean, then

$$E\left\{\frac{1}{z}\right\} \geq \frac{1}{E\{z\}}. \tag{10.4}$$

Applying this inequality to Eq. (10.3) results in the inequality

$$\sigma_{nw}^2(K,2) \geq 1 + \frac{1}{K}. \tag{10.5}$$

We now state the following theorem:

Nonconcurrent Processing Theorem: Let an input data element in the vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{N-1}$ *have any p.d.f. Then under Assumptions 2 to 5, Section II,*

$$\sigma_{nw}^2(K,N) \geq \frac{K+1}{K-N+2}. \tag{10.6}$$

Proof: The proof of this theorem is by induction and again closely follows the proof of the General Moment Theorem. We have shown that it is true for $N = 2$ (Eq. (10.5)). Thus the theorem is true for all integers less than or equal to some upper bound: $N - 1$. We show this is true for any N, which implies that it is true for any $N \geq 2$.

Again the $GS_{K,N}$ processor is decomposed as shown in Fig. 8 and further reduced as shown in Fig. 9. Neither this decomposition nor reduction depend on the p.d.f.'s of the input data. Also the concurrent data entering the $GS_{K-1,N-1}$ processor satisfy Assumptions 2 to 6. In addition, the non-concurrent data conditioned on $\mathbf{x}_{N-1}$ and X_{N-1} satisfy Assumptions 2 to 6.

Thus if $\hat{\sigma}_{K-1,N-1}^2$ is the transient unnormalized output noise power residue using nonconcurrent processing of the $GS_{K-1,N-1}$ canceller where $\hat{\sigma}_{K,N}^2 = \hat{\sigma}_{K-1,N-1}^2$ and if the Nonconcurrent Processing Theorem is true for $N - 1$, then

$$\frac{E\{\hat{\sigma}_{K-1,N-1}^2 \mid \mathbf{x}_{N-1}, X_{N-1}\}}{\lim\limits_{K\to\infty} E\{\hat{\sigma}_{K-1,N-1}^2 \mid \mathbf{x}_{N-1}, X_{N-1}\}} \geq \frac{K}{K-N+2}. \tag{10.7}$$

Again note that in the limit taken above, K goes to infinity only in the $GS_{K-1,N-1}$ structure and not in the first-level processor that precedes the $GS_{K-1,N-1}$ structure.

From Eq. (10.7), it follows that

$$E\{\hat{\sigma}_{K,N}^2\} \geq \frac{K}{K-N+2} \cdot \lim_{K\to\infty} E\{\hat{\sigma}_{K-1,N-1}^2\}. \tag{10.8}$$

We know that

$$\lim_{K \to \infty} E\{\hat{\sigma}^2_{K-1,N-1}\} = \sigma^2_{nw}(K,2) \geq 1 + \frac{1}{K}. \tag{10.9}$$

Substituting Eq. (10.9) into Eq. (10.8) results in the theorem being proved.

Note that if the input noises are Gaussian, then the lower bound given by Eq. (10.6) is almost achieved. Hence, this assumption results in almost the "best case" performance.

XI. SLIDING WINDOW GS CANCELLER CONVERGENCE

The sliding window GS canceller differs from the block processing schemes previously described in that data are processed on a sample point by sample point basis rather than in complete blocks. For the sliding window GS canceller, the GS weights are estimated every time step. Thus, the sliding window canceller tends to adapt to a nonstationary noise environment better than a block processor canceller does. The weights that are calculated in the GS structure are based on a fixed number of past samples from a given point in time. When the sampler steps one time interval, the newest sample is included in this estimate and the oldest sample is discarded.

The sliding window GS canceller is described mathematically as follows. For any indexed time instant j, set

$$x_0(j) = [x_m(j), x_m(j-1), \ldots x_m(j-K+1)]^T,$$

$$\mathbf{x}_n(j) = [x_n(j), x_n(j-1), \ldots, x_n(j-K+1)]^T, \qquad n = 1,2,\ldots,N-1 \tag{11.1}$$

where K is the number of samples used to calculate each of the GS weights. From Fig. 1(a), the input vectors into the mth level of the GS structure are defined as

$$\mathbf{x}_n^{(m)}(j) = [x_n^{(m)}(j), x_n^{(m)}(j-1), \ldots, x_n^{(m)}(j-K+1)]^T, \tag{11.2}$$

where $\mathbf{x}_n^{(1)}(j) = \mathbf{x}_n(j)$ and $x_n^{(1)}(j) = x_n(j)$. The outputs of the two-input GSs at the mth level are given by

$$x_n^{(m+1)}(j) = x_n^{(m)}(j) - w_n^{(m)}(j)\, x_{N-m}^{(m)}(j), \qquad \begin{array}{l} n = 0,1,\ldots,N-m-1 \\ m = 1,2,\ldots,N-1 \end{array} \tag{11.3}$$

where

$$w_n^{(m)}(j) = \frac{\mathbf{x}_{N-m}^{(m)t}(j)\, \mathbf{x}_n^{(m)}(j)}{\mathbf{x}_{N-m}^{(m)t}(j)\, \mathbf{x}_{N-m}^{(m)}(j)}. \tag{11.4}$$

We state and prove the following theorem:

Sliding Window GS Canceller Theorem: If the input samples are identically distributed r.v.'s, then the p.d.f.'s of a given transient performance measure (output noise power residue or signal-to-noise ratio) of the sliding window GS canceller and the concurrent GS canceller are identical. Hence all average transient performance measures are equal for the two processing schemes.

Proof: The proof of the above theorem is rather simple. Firstly, the output sequence of the sliding window GS canceller is identical to that of an output sequence generated by using successive concurrent block processing and retaining only the last output data sample. To see this, let the input data vectors to the concurrent block processor at time j be defined by Eq. (11.1). Let the K-length output sequence of this concurrent GS canceller be $z_1, z_2, \ldots, z_K$. Now z_K is exactly equal to the output of the sliding window GS canceller at time instant j; hence, the equivalence of the two processing schemes.

Secondly, the p.d.f.'s of a given transient performance measure for the concurrent block processor are identical for each z_k, $k = 1, 2, \ldots, K$. This results because all inputs are identically distributed and the computed canceller weights are independent of the ordering (or permutation) of the samples in the K-length input vectors as long as the permutation is the same for all N, K-length input vectors. Hence, the p.d.f. of any transient performance measure is identical for the two types of cancellers and the theorem follows.

XII. OVERMATCHING DEGREES OF FREEDOM

If we examine the expression for $\mathrm{SNR}_{nw}(K,N)$ given by Eq. (8.6) for the nonconcurrent GS canceller, we find that for a fixed number of input samples K, the $\mathrm{SNR}_{nw}(K,N)$ decreases monotonically as the order of the GS structure N. Hence, for a given input noise scenario, increasing the order of the GS canceller may lead to a noisier output (this phenomenon also occurs for concurrent processing).

To illustrate this problem, let us say that for a variety of input noise scenarios a $\mathrm{GS}_{K,N}$ canceller yields good cancellation performance and is therefore specified in the design. However, suppose for a specific noise scenario that only an Lth order GS canceller is needed for good performance where $L < N$. Hence, for this specific scenario there is a loss of cancellation performance by using a $\mathrm{GS}_{K,N}$ canceller instead of a $\mathrm{GS}_{K,L}$ canceller. For this case, what occurs is that at the $(L - 1)$th level of cancellation in the $\mathrm{GS}_{K,N}$ canceller, the input noises in the main channel are essentially cancelled. Thereafter, in each succeeding level the noise residue only increases. This phenomenon is called "overmatching the degrees of freedom (DOF)." For optimality, we should have stopped the cancellation process after the $(L - 1)$th level. Note that the number of DOFs for a $\mathrm{GS}_{K,N}$ canceller is $N - 1$.

By using the expression given for $\mathrm{SNR}_{nw}(K,N)$ in Eq. (8.6), we can quantify this loss for nonconcurrent processing. This loss is given by

$$\mathrm{LOSS}_{nw} = \frac{\mathrm{SNR}_{nw}(K,L)}{\mathrm{SNR}_{nw}(K,N)} = \frac{K - L + 2}{K - N + 2}. \tag{12.1}$$

One method suggested by Lewis and Kretschmer [9] for overcoming the effects of overmatching the DOFs is to monitor the nonconcurrent noise powers at each level of the main channel in the GS structure. The main channel is terminated in the GS structure where the noise power is a minimum. This point in the GS structure varies as the noise environment changes or equivalently as more or fewer DOFs are needed.

XIII. SUMMARY

The open-loop GS canceller is shown to be numerically identical with the Sampled Matrix Inversion (SMI) algorithm in the transient state if infinite numerical accuracy is assumed. Three forms of

the GS canceller are discussed and analyzed—concurrent, nonconcurrent, and sliding window processing. Previous convergence results for concurrent and nonconcurrent SMI cancellers that assume Gaussian inputs have been reproduced by using the GS structures as an analysis tool. In addition, new results are obtained for when the input noises are not Gaussian. Furthermore, the sliding window GS canceller is shown to have the same convergence properties as the concurrent GS canceller. The deleterious effect of "overmatching the degrees of freedom" is discussed.

XIV. REFERENCES

1. B.L. Lewis and F.F. Kretschmer, Jr., NRL, published works of limited distribution, dating back to Feb. 1974.

2. R.A. Monzingo and T.W. Miller, *Introduction to Adaptive Arrays* (John Wiley and Sons, New York, 1980), Ch. 8.

3. W.F. Gabriel, "Building Block for an Orthonormal-Lattice-Filter Adaptive Network," NRL Report 8409, July 1980.

4. M.A. Alam, "Orthonormal Lattice Filter—A Multistage, Multichannel Estimation Technique," *Geophys.* **43**, 1368-1383 (1978).

5. B. Friedlander, "Lattice Filters for Adaptive Processing," *Proc. IEEE* **70**(8), 829-867 (1982).

6. Karl Gerlach, "Fast Orthogonalization Networks," *IEEE Trans.* **AP-34**(3), 458-462 (1986).

7. I.S. Reed, J.D. Mallett, and L.E. Brennan, "Rapid Convergence Rate in Adaptive Arrays," *IEEE Trans.* **AES-10**, 853-863 (1974).

8. L.E. Brennan and I.S. Reed, "Digital Adaptive Arrays with Weights Computed from and Applied to the Same Sample Set," Proceedings of the 1980 Adaptive Antenna Symposium, RADC-TR-80-378, Vol. 1, Dec. 1980.

9. B.L. Lewis and F.F. Kretschmer, Jr., NRL, private correspondence.

Appendix A

PROBABILITY DENSITY FUNCTION ASSOCIATED WITH A TWO-INPUT GS CANCELLER

In this appendix, we outline a derivation for obtaining the p.d.f. of the transient noise power residue and the transient SNR of a two-input GS. For this analysis the estimated weight is applied to a data set that is independent of the data set that calculated the weight (nonconcurrent processing).

Let $\mathbf{x}_0$ and $\mathbf{x}_1$ be the K-length input vectors associated with the main and auxiliary channels respectively. The estimated weight of the GS canceller is then given by

$$\hat{w} = \frac{\mathbf{x}_1^t \mathbf{x}_0}{\mathbf{x}_1^t \mathbf{x}_1}. \tag{A1}$$

For nonconcurrent processing the transient noise power residue is given by

$$\hat{\sigma}^2 = 1 + |\hat{w}|^2. \tag{A2}$$

We derive the p.d.f. of $|\hat{w}|^2$, from which the p.d.f. of $\hat{\sigma}^2$ is easily attainable.

Now let $z = |\hat{w}|^2$ or

$$z = \frac{|\mathbf{x}_1^t \mathbf{x}_0|^2}{(\mathbf{x}_1^t \mathbf{x}_1)^2} = \mathbf{x}_0^t \frac{\mathbf{x}_1 \mathbf{x}_1^t}{\mathbf{x}_1^t \mathbf{x}_1} \mathbf{x}_0 \frac{1}{\mathbf{x}_1^t \mathbf{x}_1}. \tag{A3}$$

Let

$$v = \mathbf{x}_0^t \frac{\mathbf{x}_1 \mathbf{x}_1^t}{\mathbf{x}_1^t \mathbf{x}_1} \mathbf{x}_0. \tag{A4}$$

We derive $p(v \mid \mathbf{x}_1)$. The matrix $\mathbf{x}_1 \mathbf{x}_1^t / \mathbf{x}_1^t \mathbf{x}_1$ can be written as

$$\frac{\mathbf{x}_1 \mathbf{x}_1^t}{\mathbf{x}_1^t \mathbf{x}_1} = \Phi^t \, \Lambda \, \Phi, \tag{A5}$$

where Φ is a $K \times K$ unitary matrix and Λ is the diagonal matrix of eigenvalues where the first diagonal element equals 1 and all others equal 0. Thus

$$v = \mathbf{x}_0^t \Phi^t \Lambda \Phi \mathbf{x}_0 = (\Phi \mathbf{x}_0)^t \, \Lambda \, \Phi \, \mathbf{x}_0. \tag{A6}$$

Set a new K-length vector $\mathbf{y} = \Phi \mathbf{x}_0$. It is easy to show that if $\mathbf{x}_0$ is a normalized K-length multivariate complex circular Gaussian vector, then $\mathbf{y}$ is the same. If y_i is the first element of $\mathbf{y}$, then from Eq. (A6)

$$v = |y_1|^2. \tag{A7}$$

The p.d.f. of $|y_1|^2$ is the well-known second order chi-square and thus it can be shown that

$$p(v \mid \mathbf{x}_1) = e^{-v}, \quad v > 0. \tag{A8}$$

Note that this p.d.f. is independent of x_1. As a result,

$$p(v) = e^{-v}, \qquad v > 0. \tag{A9}$$

From Eq. (A3) and elementary probability theory we can show that

$$p(z \mid x_1) = \mathbf{x}_1^t \mathbf{x}_1 e^{-z \mathbf{x}_1^t \mathbf{x}_1}, \quad z > 0$$

$$= u e^{-zu} \quad , \quad u \leq 0 \tag{A10}$$

where $u = \mathbf{x}_1^t \mathbf{x}_1$. It is known that the p.d.f of u is given by

$$p(u) = \frac{1}{(K-1)!} u^{K-1} e^{-u}, \quad u \geq 0. \tag{A11}$$

Hence

$$p(z) = \int_0^\infty p(x \mid u)\, p(u)\, du \tag{A12}$$

$$= \int_0^\infty \frac{1}{(K-1)!} u^K u^{-(z+1)u}\, du. \tag{A13}$$

It is elementary to show that the above integral reduces to

$$p(z) = \frac{K}{(z+1)^{K+1}}, \quad z \geq 0. \tag{A14}$$

If we set $\eta = \hat{\sigma}^2$, then it follows from Eqs. (A14) and (A2) that

$$p(\eta) = \frac{K}{\eta^{k+1}}, \quad \eta \geq 1. \tag{A15}$$

Appendix B

FIRST MOMENT BOUND

If z is a random variable with $z \geq 0$, $0 < E\{z\} < \infty$ and $E\{1/z\} < \infty$, then

$$E\left\{\frac{1}{z}\right\} \geq \frac{1}{E\{z\}}. \tag{B1}$$

We use the Cauchy-Schwarz inequality to show this. Let $p(z)$ be the p.d.f. of z. Now

$$\int_0^\infty p(z)\,dz = \int_0^\infty \sqrt{\frac{p(z)}{z}} \cdot \sqrt{zp(z)}\,dz = 1, \tag{B2}$$

where the square root function shown above is the positive square root function. Using the Cauchy-Schwarz inequality,

$$\int_0^\infty \frac{p(z)}{z}\,dz \cdot \int_0^\infty zp(z)\,dz \geq \int_0^\infty \sqrt{\frac{p(z)}{z}} \cdot \sqrt{zp(z)}\,dz. \tag{B3}$$

Thus

$$\int_0^\infty \frac{p(z)}{z}\,dz \cdot \int_0^\infty zp(z)\,dz \geq 1. \tag{B4}$$

Equation (B1) follows from Eq. (B4).